Realism in Mind

A dissertation submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Philosophy

Ricardo Restrepo Echavarría

University of Canterbury
2010

For two surrealists,
Inés and Marce

Acknowledgments

Thanks to my supervisors Graham Macdonald and Cynthia Macdonald for the discussion, the blend of wide support, encouragement, and skepticism. The mix of independence and discipline that they encouraged (or is it “tolerated”?) was essential. In midst of it all, because of them I got to be in Belfast in time for the event where Gerry Adams and Ian Paisley put the guns downs and sat at the same table.

Thanks to my other supervisor Jack Copeland for discussion.

Thanks to the Marsden Fund that the Macdonalds were awarded and with which this research was funded.

Thanks to Julia Capon for co-juggling the adventure that was our film, which in turn made possible completing this thesis.

Thanks to Marce for introducing me to the Land of the Long White Cloud and the University of Canterbury.

Abstract- The thesis develops solutions to two main problems for mental realism. Mental realism is the theory that mental properties, events, and objects exist, with their own set of characters and causal powers. The first problem comes from the philosophy of science, where Psillos proposes a notion of scientific realism that contradicts mental realism, and consequently, if one is to be a scientific realist in the way Psillos recommends, one must reject mental realism. I propose adaptations to the conception of scientific realism to make it compatible with mental realism. In the process, the thesis defends computational cognitive science from a compelling argument Searle can be seen to endorse but has not put forth in an organized logical manner. A new conception of scientific realism emerges out of this inquiry, integrating the mental into the rest of nature. The second problem for mental realism arises out of non-reductive physicalism- the view that higher-level properties, and in particular mental properties, are irreducible, physically realized, and that physical properties are sufficient non-overdetermining causes of any effect. Kim's Problem of Causal Exclusion aims to show that the mental, if unreduced, does no causal work. Consequently, given that we should not believe in the existence of properties that do not participate in causation, we would be forced to drop mental realism. A solution is needed. The thesis examines various positions relevant to the debate. Several doctrines of physicalism are explored, rejected, and one is proposed; the thesis shows the way in which Kim's reductionist position has been constantly inconsistent throughout the years of debate; the thesis argues that trope theory does not compete with a universalist conception of properties to provide a solution; and shows weakness in the Macdonald's non-reductive monist position and Pereboom's constitutional coincidence account of mental causation. The thesis suggests that either the premises of Kim's argument are consistent, and consequently his reductio is logically invalid, or at least one of the premises is false, and therefore the argument is not sound. Consequently, the Problem of Causal Exclusion that Kim claims emerges out of non-reductive physicalism does not force us to reject mental realism. Mental realism lives on.

TABLE OF CONTENTS

Introducing What's to Come

This book is divided into two main sections. Each section exposes a prominent threat to Mental Realism and provides arguments to defeat it. Mental Realism is the doctrine that "mental properties are real properties of objects and events; they are not merely useful aids in making predictions or fictitious manners of speech" (Kim 1993a, 344). To be a Mental Realist "your mental properties must be *causal properties*-properties in virtue of which an event enters into causal relations it would not have otherwise have entered into" (Kim 1989b, 279). More broadly, let us say that Mental Realism is the doctrine that mental properties, objects, and events are a part of the causal structure of the world, and that they enter into causal relations as mental. The addition of "as mental" seems to be trivial. Could mental properties contribute to causal processes other than as mental properties? Could mental events or mental objects enter into causal transactions as other than what they are- mental? However, because some argue that mental properties are reductively identical to physical properties actually doing all the causal work, the view that mental properties have an active place in the production of behaviour is deflated to be the view that mental properties have an active place in the production of behaviour, not insofar as they are mental, but only insofar as they are non-mental. However, when we explain behavior we seem to be talking about certain mental events going on in the world, where agents have certain mental properties, *in virtue of which* their behavior comes about. The recognition that this is so, even in the last analysis, is the view that mental properties are irreducibly mental and causal.

A presupposition of Mental Realism is that the core assumption of mentalistic explanation, that mental properties are exemplified by agents in events (at times; or by events at times), is not undermined by our further assumptions as Scientific Realists and physicalists.

Stathis Psillos has defended and stated Scientific Realism in terms that (perhaps inadvertently) negate the reality of mind. Thus, Scientific Realism becomes a threat to Mental Realism. As formulated by Psillos, Scientific Realism undermines the essential role for minds in the construction of theories, the bearers of scientific truth-value itself, for example. It also undermines the idea that psychology is a scientific field whose subject matter is the mental. The first section will take note of the implausibly big cost of holding on to such a statement of Scientific Realism and propose corresponding adaptations to the theory. This section will conclude with an extended defense of computational cognitive science from an argument that has received little attention but is immanent in John Searle's work and directly relates the realist philosophy of science with the philosophy of mind.

Jaegwon Kim has generated a problem for mental realists who believe that all mental properties are physically realized. The problem I have in mind is the Causal Exclusion Problem. This problem is the subject of the second section and may be stated as follows. The realization of physical properties is what the realization of all other properties depends on. Given that the physical world is the world of concrete reality, the realization of physical properties is, in this sense, ontologically basic. Because in the last analysis everything in nature ultimately depends on these ontologically basic things, any property that is not a physical property is either epiphenomenal- making no causal contribution- or identical with these basic things. If mental properties are epiphenomenal, then because we ought not to believe in things that are not part of the causal structure of the world (Alexander's Dictum), we ought not to believe that mental properties exist. In that case Mental Realism is undermined. If what we took to be mental properties are reduced to physical properties (at least under a certain conception of physical properties) we defeat Mental Realism by admitting that in the last analysis what we took to be mental properties figuring in the causal origins of action were really physical properties. The second section will consider the most prominent ways of dealing with this problem and suggest another way out of this threat to Mental Realism.

The writing is divided into two main sections, which are in turn divided into chapters, which are in turn divided into sub-sections. The numbering for the headings follows that order.

1. *Scientific Realism in Psychology: Necessary Adaptations*

Stathis Psillos builds a very compelling case for Scientific Realism and against its alternatives in his *Scientific Realism: How Science Tracks Truth*. I am not, however, concerned primarily with the justification of Scientific Realism.[1] It is a plausible thesis, or set of theses, that has in general been convincingly argued for by Psillos and others. Thus, the only justification for Scientific Realism that will be provided here will be insofar as it is necessary to account for the truth of Mental Realism.

What concerns us in this section is that as Psillos formulates it, Scientific Realism rules out Mental Realism. Mental Realism asserts the existence of mental reality, mental properties, mental events, and mental substances. It may come as a surprise that Mental Realism asserts the existence of mental substances without having Cartesian Dualism as a consequence. However, this surprise is quickly quelled when we consider that a mental substance *just is* a substance with mental properties, in our case, selves or minds. That mental substances are immaterial – or that they have existence separate from the physical- is another thesis for which more premises than endorsed here are required. It is interesting that Scientific Realism has been formulated in such a way that it rules out psychology. Perhaps it is a reactionary historical move against the idealist contenders to which Scientific Realism posed a challenge and an alternative. But on the very grounds of a core Scientific Realistic conception, that we reliably find realities through scientific investigation, we should not rule out our finding

[1] In the present discussion what is within the realm of science is narrowly construed to be nature; we ignore pure mathematics.

mental reality through scientific investigation. Further, our own psychology is an essential instrument of ours in the very investigation not only of mental reality, but also of non-mental reality. This is a section that adapts standard Scientific Realist theses to make space for the mental on a par with everything else Scientific Realism incorporates. It does this by revising the standard Scientific Realist commitments, and adapting and integrating some prominent avenues to Scientific Realism in order to cohere with Mental Realism. According to Psillos, Scientific Realism is constituted by the following three standard theses.

1. The Metaphysical Thesis: The world has a definite and mind-independent natural-kind structure.
2. The Semantic Thesis: Scientific theories should be taken at face value, seeing them as truth-conditioned descriptions of their intended domain, both observable and unobservable. Theoretical assertions are not reducible to claims about the behaviour of observables, nor are they merely instrumental devices for establishing connections between observables. The theoretical terms featuring in theories have putative factual reference. So, if scientific theories are true, the unobservable entities they posit populate the world.
3. The Epistemic Thesis: Mature and predictively successful scientific theories are well-confirmed and approximately true of the world. So, entities posited by them, or at any rate, entities very similar to those posited, do inhabit the world (Psillos 1999, xix).

It is no wonder that scientific realists are by and large concerned with the status of physics. The discipline theorizes about what may most plausibly be cast as mind-independent, at least at first glance;[2] the meaning of the scientific assertions has been most scrutinized by people, the positivists and to some extent today's philosophical players, whose scientific background is dominated by physics; physics is also the most mature of the scientific fields, and is the most basic of the sciences. In a way the metaphysical physicalist consensus amongst a wide cut of contemporary philosophers is an expression of the commitment to physics as the best established rigorous domains of fields of knowledge. In this section I will try to address how we can be Scientific Realists about the mind, and in so doing I will note conceptual reforms required to meet considerations in conflict with Psillos's theory of Scientific Realism. This is in line with the ideal of a unified conception of nature and the mind's place within it.

1.1 The Metaphysical Thesis: The Mind-Independent Study of What?

In this chapter I would like to discuss the way in which reference and truth are mind-independent. And the way in which they are not. I share Psillos's broadly naturalist approach, but there are tensions within the Scientific Realist picture he advocates, as well as tensions arising from consideration external to the issues explicitly treated in his book. In particular, with respect to the second kind of tensions, I am concerned primarily with issues that arise from a realist stance towards psychology. In this chapter I want to show some of the ways in which such tensions arise and provide some adaptations that need to be made to the Metaphysical Thesis.

The Subject of the Scientific Field

Trivially, it would seem, but in contradiction to Psillos's stated thesis, if the task of psychology is to tell us the nature of mind, as is plausible, then psychology is not

[2] Though see Stairs (1991).

mind-independent. It centrally involves the mind. A first approximation to overcoming this difficulty is noting that the nature of mind is independent of what we theorize about it. To be realists about mind we must assert that its nature is not determined by our theorizing about it, denying that our theorizing structures minds, which would otherwise have no structure or have a different structure.[3] It is very plausible that a developed biopsychology will tell us about at least some of the fundamental features of mind. Broadly speaking, two of the prominent approaches to mind are ones that come from biology and the other from computer science. If the theory of evolution is true, then we have a framework in place already to understand some of the fundamental features of mind. According to this approach, the mind is an evolved organ comprised of particular evolved features. The work of evolution is mind-independent and the evolved features of mind conferred by natural selection and other evolutionary factors on existing mental creatures like ourselves are in place prior to our doing anything about them, and *in this way* are mind-independent. Another prominent approach to mind is computational, founded on the ideas of Alan Turing (1936, 1950). According to this approach, the mind is an information processing machine, whose fundamental feature is performing computations over symbols according to a program. *Which computations* our minds implement are *in this way* mind-independent. Both approaches have mind at their centre because it is their subject of inquiry, but both approaches also make use of theoretical backgrounds that we understand enough and which have been adequate in several successful research areas involving non-mental aspects of the world. Further, that the particular predictions that we gather from such theoretical models can

[3] The use of “structure” here is not the one relevant to Russell that is treated later. Rather, it is more akin to having a character.

be incorrect shows that they have the capacity not only to *be* false, but also *shown* to be false if they are. This indicates not only that the theories we use to understand mind have enough contact with nature to not be trivially true, but also they have a real risk of being made and shown false. In other words, nature is doing work in regulating our theoretically guided practices and this is just what we want of anything that might be claimed to be knowledge of a natural phenomenon.

Fair Play For Humans

There are also features of mind that we either as a social group or as individuals "make." For example, we may construct certain games and come to employ certain mental strategies for succeeding at playing the game we designed through reflection and training. In these respects also, we find fields of research that aim to discover their nature: social psychology, social neuroscience, and economic psychology are some examples. On such occasions it is plausible that at least some of such features of mind are artifacts. There are some who believe that the way we attain knowledge of artifacts is radically different from knowledge of kinds whose features are not so particularly dependent on human intention and craft. But the grounds for this are not well-supported.

Artifacts are mind-dependent in a particular way: *our* uses and designs are constitutive of their nature. Schwartz argues that this feature of artifacts indicates that artifacts are not referred to in the same way as kinds whose existence is independent of human design and use.[4] In other words, the way we attain epistemic access to artifacts, the way we come to know them, is radically different from the way we attain epistemic

[4] Kornblith (2007) quotes Schwartz (1978) Schwartz (1980) and Schwartz (1983).

access to natural kinds. However, we can refer to artifacts, like chairs and doorstoppers, in the same way we refer to individuals and natural kinds with their corresponding terms, like “Gerry Adams” and “water,” without knowing much about them, and judge things of them that are true or *false* (Kornblith 1980; 2007). For example, we may refer to rheostats without knowing what distinguishes them from plugs, batteries, or transistors. But we may nevertheless use “rheostats” to refer to rheostats, just as Hilary Putnam once pointed out the Problem of Ignorance with the example that he may refer to beeches with his use of “beeches” without knowing them apart from elms. Further, there are plenty of things we might be wrong about with artifacts. Suppose Joe points at the rheostat in front of him and says that “rheostats generate electricity.” Then Joe has said something false, even if rheostats are artifacts and this is the description he associates with the term “rheostats.” In this respect, then, knowledge claims about artifacts are subject also to the Problem of Error, as are knowledge claims of individuals and natural kinds. What the Problem of Error and the Problem of Ignorance indicate is that claims of knowledge about artifacts are on a par with claims of knowledge in any other field. As long as minds, like atoms and chairs, are able to be referred to, they can make our statements about them true or false. If in some respects the mind is an artifact, there is no reason, at least on that account, that there is a lack of objectivity in how the truth-value of our assertions about it is determined.

To sum up, minds are the truth-value-makers (or just “truth-makers”) of psychology, and as truth-makers they are independent of our theories. What can be said thus far is that (1) without minds psychology would not have a non-empty set of entities to study. That is, minds are required for the science of psychology to have successful reference. In this respect, minds are essentially involved in the very possibility of successful psychological research. And (2) minds have (at least some) properties independent of the research that investigates them.

Theories and assertions are the bearers of truth-value, but they are also things we make. So in this other way, truths about psychology are not independent of our mental practices, again apparently in contradiction with how Psillos's expresses the Metaphysical commitment of Scientific Realism.

The fact that theories and assertions are applied in all scientific fields of knowledge means that the above way in which truth is mind-dependent also applies in physics. But this should be no problem. As Richard Boyd (2006) points out, we should have fair play for humans. Beavers make dams, and in that sense some dams are dependent on the practices of beavers. But we don't consider them unreal because of that. No beavers, no (beaver-made) dams. No epistemically-governed minds making theories, like ours, no theories, and therefore no true theories. But this in no way jeopardizes that the world makes our empirical theories and assertions true or false, and it doesn't threaten the idea that those truth-makers might have existed independently of the truths they make true. Truths refer to the things that make them true, and our theories can really be true even if we created them and even if we created the things that make them true, as in the case of artifacts.

Feedback Effects Can Be Deceptive but Overcome

In psychology, however, our inquiry and the subject of our inquiry seem to have some particular features which may lead some to doubt that psychology can really reveal properties with enough independence to be objects of scientific knowledge. Two ways one might worry are as follows: We might worry that the very procedure of engaging in research changes the very nature of the phenomenon under consideration. To be more specific, we might worry that the results of research influence the very structure of the minds of participants in psychological experiments.

To be sure, participants who live in a society where the results of psychological experiments are put to some use are subject to the influences of those uses and may walk into the lab under those influences. For example, it is a well-confirmed fact that people tend to pick products placed on their right hand side and which are blue. Thus, marketers fight to exploit this human bias by choosing blue packaging and placing their products, like say detergent, on the right hand side of the supermarket shelves. Participants may then have clothing that is not so clean because they wash their clothing with detergent that is not necessarily particularly good, and then walk into the lab to engage in the experimenter's design. In this case we see that there is a perhaps systematic occurrence of certain economic choices which result in the effect that the participants walk into the lab with. Suppose for definiteness, that the experimenter's design is made to test for properties of visual depth-detection. Then, it seems that, provided that wearing dirty clothes does not have effects on every cognitive property of the clothes wearer, it is safe to say that it won't have an effect on her visual depth-detection. But perhaps it does. Does that mean that visual depth-detection is not able to be investigated? No. For you can fix the experimental design to insulate the properties under investigation from the effects of the participants' choice of detergent; and you can change what we take the experiment to be evidence for.

Self-Reference, Stich's Paradox and Epistemic Immunity

The second way one might worry that psychology cannot be an objective science is that there is a self-reference involved in the study of mind that seems to not exist in the study of other aspects of nature. This self-reference might tempt some to be suspicious of what is revealed through psychological experiment because we are examining psychological mechanisms through the employment of those very kinds of psychological mechanisms. We might conclude, with what seems to be a consequence of Psillos's Metaphysical Thesis, that inquiry into mind is positively not acceptable given Scientific Realism precisely because such inquiry presupposes the existence mind-dependent things.

But self-reference should not discourage us from being able to inquire and test nature to ascertain the truth or falsity of claims about the mind, of which we ourselves are tokens. On a softening-up note we might notice that we are made, amongst other things, of Carbon, that the investigation of Carbon itself employs the movement of Carbon in the investigator's body, and that this is not a particular problem for the investigation of this chemical element.

It is significant to point out that in fact we find a source of epistemic immunity in the study of mind that we do not find in the study of physics originating from this particularity, for we can determine the truth of some thoughts just by having certain thoughts. For example, we determine the truth of the thought that *this thought is a* *thought*, just by having it. Further, we may ascertain that we are not *irremediably* deluded by certain engrained epistemic biases, for which we have evidence through scientific inquiry, as Stephen Stich may be interpreted to conclude from experiments on human bias. If we were irremediably biased towards epistemically irrational beliefs, i.e. that we are wholly and irremediably epistemically irrational, this would undermine the very possibility of the putatively epistemically rational belief Stich argues we should adopt. This putative situation is Stich's Paradox.

Before entering into further analysis, two notes on Stich's contention are in order. Strictly speaking, Stich does not clearly commit himself to the view that all our beliefs are epistemically irrational. Stich's central thesis is that "philosophical arguments aimed at showing irrationality cannot be experimentally demonstrated are mistaken" (Stich 1984, 337). I agree with this thesis if it means that philosophical arguments cannot show that good evidence for the idea that humans are irrational in various systematic ways cannot be ascertained. That is, I agree if it means that strong evidence can be obtained for the thesis that we display epistemically irrational thought and behaviour patterns. I disagree if it means that philosophical argument cannot show that irrationality cannot cover all our truth-valuable mental life. I try to show that it can. The issue is worth resolving independent of Stich's intended interpretation, since refuting the second interpretation puts a definite barrier on how widely the irrationality contended in the first interpretation can be reasonably envisioned to cover our belief system. When this interpretation is presented as Stich's, it is in the loose sense, that this is a view worth discussing emanating from his work.

Secondly, there will be a strong link assumed between rationality and truth. Some do not think this sort of connection applies. Popper and Musgrave do not think this connection applies (Musgrave 2004), for example, and people like them can be merely entertained by the present exercise and see what can be done by assuming the link. I will assume that from an epistemic point of view, the issue of rationality is a belief-system's connection to truth. Stich displays this assumption when he detects irrationality in participants of psychological experiments particularly when they end up believing falsehoods and use false inference rules. Inference rules are truth-valuable because they can be given the form, "if P, then Q." There is also a case to be made that even if certain false beliefs can be said to be rational, like Newton's belief that physical space is flat, once the evidence was in, it became irrational to believe the same. This resonates with what could be a variety of internalism about reason, which says that what an agent has reason to believe in or do is determined by her internal intentional states. It does seem strange to say that at least many falsehoods, however, no matter if the available evidence supports it, are rational to believe in. One might be said to be irrational if one invested in a bad investment, even though it made sense with respect to the limited amount of considered information. After losing money this way, one will say "I *should* not have invested that way," acknowledging one acted with an element of irrationality. So it looks as though the investor acted internally rationally but externally irrationally. It is also partly because the belief that physical space is flat became *demonstrably false,* once the evidence for the alternative accumulated, that this belief became irrational. A kind of externalist about reasons, one which would hold that facts external to the agent determine what an agent has reason to believe in or do-independent of her current internal state of mind- will find this attractive. Two forms of epistemic rationality may be discerned to accommodate both positions: component and net. At the time, Newton's belief in flat physical space was epistemically component-ly rational, but net-ly irrational, as was the investor's actions. This allows us to say that it was component-ly rational for Newton to believe that space is flat. But the scheme also recognizes that there is something irrational in believing any falsehood, such as Newton's belief. Thus, net-ly his belief was irrational. As the evidence came in, Newton's belief became component-ly irrational, as well as net-ly irrational. That truths are regarded as rationally believed in is treated as uncontroversial in this context- they do become controversial when issues of complexity, cluttering, and when truths we do not care to know are at issue. They are not relevant here since, from an epistemological point of view, on any matter at hand, we wish to believe truths and not believe falsehoods.

Likewise, there should be a sense in which believing falsehoods is irrational, no matter what evidence an agent takes into account at a time; and, on the flip side, there should be a sense in which believing truths is rational, no matter what evidence the agent happens to have. My basic instinct for this thought is that in a scenario where a believer of a falsehood *P* which is based on ignorance of, or false beliefs about, certain evidential truths, I think the believer nevertheless has, in a certain sense, a reason, unbeknownst to him or her at a time, to believe *Not P*, a reason she would *find* if s/he were put in a better epistemic position. On the flip side, it would seem that I can hold true beliefs irrationally – on bad evidence. Yes, it is just that in my scheme I would be component-ly irrational, but net-ly rational. This is because, likewise, if I were put in a better epistemic position, I would be able to fix my evidence and *find* the good grounds for my true belief- grounds that exist independent of my current internal state of mind.

In what follows, when I use “rational” and “irrational” it will be in the sense of *net* epistemic rationality- directly connected to truth and falsity, respectively.

Stich tries to sever the link between rationality and truth. Rather, he contends, rationality must relate a cognitive system to “consequences that people actually value” (Stich 1990, 395). In agreement with Stich, pragmatic factors determine whether a true belief is valued. Pragmatic factors determine what true beliefs are relevant for the agent and at least potentially worth having. Further, a relevantly true belief is still a true belief and Stich has not ruled out that truth is something we actually value, and that actually valuing truth might well be something governing belief formation in a general way.

Kornblith (2002) developed a naturalistic theory of knowledge that asks just the question that Stich asks. He first takes a cue from Quine, who says:

> Naturalization of epistemology does not jettison the normative and settle for the indiscriminate description of ongoing procedures. For me normative epistemology is a branch of engineering. It is the technology of truth-seeking…There is no question here of ultimate value…The normative here, as elsewhere in engineering, becomes descriptive when the terminal parameter is expressed.[5]

This is a very attractive starting point for answering the question “who cares about true belief?” But it suffers from the threat of parochialism. It is desirable for epistemic rationality to apply to all, not just to those that contingently seek truth.

In many contexts justification about why we have reason to act in a certain way can stop at the invocation of a proper desire. It is therefore worth considering whether

[5] Quine (1986) quoted by Kornblith (2002: 138).

epistemic norms might themselves be grounded in desire. [6] What is required is a general point about the value of truth, avoiding the putative consequence that epistemic norms are grounded on some particular idiosyncratic desire; for we want epistemic norms to apply universally and we do not share idiosyncratic desires. Kornblith proposes that a truth-seeking norm- or in present terms, epistemic rationality- applies to all organisms that desire anything at all. This is because an organism that does not have a truth-seeking norm that applies to its cognitive and affective functioning truncates its own desires. One way in which this would be done, is by making the satisfaction of particular desires quite unlikely. Thus, according to this applicable norm, agents have reason to act in accordance with it. Consider an organism that desired pizza. Unless the cognitive system of the organism generates truths about pizzas, where they are located for example, it will not get pizza except by luck. In that sense, an organism that desires pizza is bound by truth-seeking norms. This holds for any of the multifarious kinds of desires of humans, just so long as humans desire at all.[7] As such, truth-seeking norms, or epistemic rationality, apply not categorically, but hypothetically. But this application is not parochial since the norm applies to all organisms that desire anything at all. Truth, therefore, is something to be cared for quite generally.

Now to Stich’s Paradox. If it is rational to believe that we are inevitably believers of only irrational beliefs, as Stich contends, then it is a rational belief- given this truth- *that all we believe is irrational*. It follows from what this belief says, that this very belief is itself irrational. But, if it is an irrational belief that each belief we believe

[6] John Broome is a notable exception. He finds no reason to do as he desires, as he argued in his lectures on Rationality at the University of Canterbury Philosophy Department in New Zealand held between September and November 2005.

[7] For a more detailed account see Kornblith (2002, Chp. 5).

is irrational, then, given that net-ly irrational beliefs are false, we must believe at least one rational belief. What else could make the belief *that all we believe is irrational* irrational? But if we believe at least one rational belief, then the belief that each belief we believe is irrational is irrational, since it is false. So if the belief is rational, then it is irrational; and if it is irrational, then there is at least one rational belief, so again, it is irrational. Therefore, the belief that all we believe is irrational is irrational. The solution to Stich's Paradox seems to be that there is no rational belief saying that all we believe is irrational. That belief must itself be irrational.

While not explicitly asserting it, the proposed solution does admit that we are irrational at times. What is refuted is that we *must* be irrational, and "must" here is interpreted generally, as that we are not in a *position* where rationality is available. The rock climber can be positioned to grab a hold and yet fail to grab it, and likewise, a cognitive agent may be positioned to attain certain truths without attaining them. The more training the rock climber gets, in fact by grabbing certain holds, the better s/he will be positioned in the future to do harder routes. The more training the cognitive agent gets, in fact by believing certain truths, the better s/he will be positioned in the future to believe more interesting truths. What is refuted, in other words, is that we are not in a position to generate true beliefs, from which the option to generate further, more sophisticated true beliefs, in the future is derived. We can become better believers, if you like. That we are in such a position is presumed by the taking of empirical evidence and theory as yielding results we can trust, as Stich does.

In this juncture, the reader may be reminded of the Liar's Paradox, which asserts: *this statement is false*. If it is true, then it must be as it says it is, namely false. If it is false, then it is as it says it is, so it is true. Let us call the analogous paradox Stich's Lying Paradox. This paradox emanates from the putative belief that *all our beliefs are irrational*. It follows from this belief that *this very belief,* being one of all our beliefs, *is irrational.* But if it is irrational, then it is as it says it is, namely irrational. But if it is irrational, then it is as it says it is; so, given that this belief is true (add "and relevant" if you wish), it is rational. The belief is rational if and only if it is irrational.

The very possibility of rational belief, required by the Stich's argument based on experiments on human bias, cannot imply that we are incapable of acquiring rational beliefs about the phenomenon in question. It would be hard for Stich to deny that the process by which he arrived at his conclusion was a process he does not epistemically commend. Agreeing that there are trustworthy mechanisms, or mechanisms more trustworthy than others, for acquiring beliefs about a subject is accepting that even if we are biased in many ways we have the capability to be rational, or more rational.[8]

The argument can clearly be stated in terms of statements and truth rather than beliefs and rationality, by making the appropriate substitutions, yielding the original Liar's Paradox.

It is apparent that the considered interpretation of Stich's comments also leads to Stich's Lying Paradox, and this may be taken to indicate one of two things. One, that the result is a reduction to absurdity of the considered view, namely that we are and must be irrational. Or two, that the fate of such a view will be tied to the uncertain and unhopeful fate of "this statement is false."

[8] Kornblith (1992a) describes how we are in some ways employing faulty logic in belief formation, like following the law of small numbers, though he argues that this is a good epistemic heuristic nevertheless, and how systematically biased behavior can nevertheless be adjusted (Kornblith 1992b).

It is unhopeful because the various solutions posed to The Liar's Paradox notice a certain deficiency in the character of what is expressed. It will perhaps be useful to consider a way to analyze the Liar's Paradox that brings out what I think dissolves the paradox and is in agreement with the proposed solution to Stich's Paradox, namely to deny the existence of a rational belief that all we believe is irrational; and by extension, in Stich's Lying Paradox case, to deny the possible existence of the rational belief that this very belief is irrational.

The Liar's Paradox has been tried to be analyzed in terms of the paradox of the barber. The story goes that there is a town in which a barber lives and this barber is such that he shaves all and only those who do not shave themselves living in this town. So who shaves the barber? Given that he shaves only those who do not shave themselves, he cannot shave himself. But if he doesn't shave himself, then given that he shaves all those who don't shave themselves, he must shave himself. But if he shaves himself, then given that he doesn't shave those who shave themselves, he doesn't shave himself. So he shaves himself if and only if he doesn't shave himself. The way out of this is just to recognize that such a barber, a barber with the specified properties, is just impossible- there just is not and cannot be such a barber. So essentially, the solution to Stich's Paradox has the same structure as the solution to the paradox of the barber.

It may be apparent that this kind of solution will not apply to the Liar's Paradox, and some may argue by extension to Stich's Paradox and Stich's Lying Paradox. It is, of course, apparent that "this statement is false" is not only a possible, but an actual, statement. However, I do think that the way out the paradox of the barber in application to the Liar's Paradox has not received proper consideration. "This statement is false" has the appearance of a statement, in virtue of its standard assertoric intonation, proper grammar, and being built out of meaningful words. However, “colourless green ideas sleep furiously” also has these properties, and yet it is not a real statement. It may well be a mark of statements that they have such properties, but having them is not sufficient for a string of words to be a statement. Further, if it is an essential property of statements that they have truth-conditions, as seems plausible, then we get an explanation of why “this statement is false” and “colourless green ideas sleep furiously” are not statements- they have no truth-conditions, they cannot be true or false. There are no possible ways a world would be if “colourless green ideas sleep furiously” is true. So the same solution to the Liar’s Paradox would apply: there is no such statement, such that it is false if true, and true if false. *Mutatis Mutandis* there is no net-ly rational belief that all our beliefs are net-ly irrational, and no belief that all our beliefs are false.

In considering the Liar’s Paradox, Tarski urges that we ought not to treat it as a joke or sophistry. “It is a fact that we are here in the presence of an absurdity…If we take our work seriously, we cannot be reconciled with this fact. We must discover its cause…” (Tarski 1944, 73). While a complete theoretical treatment of the Liar’s Paradox is beyond the scope of this book, the cause proposed here- the absence of truth-conditions for the putative assertion in question- deserves further investigation. This will have to be at some other time, given that it is outside the main purpose of this dissertation. What the proposed avenue for a solution advocated here illustrates is that in tying the fate of Stich’s considered view to the Liar’s Paradox, will either undermined it completely as a thesis, or will heavily hurt it by exhibiting what Sainsbury (1995) describes as a semantic defect, or what Tarski calls “an absurdity.”

In sum, philosophical argument is relevant to the investigation of rationality; this does not exclude empirical evidence from the investigation; however, that empirical evidence cannot imply that we are completely irrational, without undermining its own epistemic force and contradicting itself.

Fair Play for Theorists and the Theory of Reference

Now let us continue with Psillo's Scientific Realism, and particularly the interpretation of scientific language. Psillos correctly believes, I think, that statements are true just in case the entities referred to stand in the referred to relations. Here I will argue that Psillos's theory of reference itself presupposes the existence of minds. Psillos's Causal-Descriptivist theory of reference has two postulates:

1. A term t refers to an entity x if and only if x satisfies the core causal description associated with t.

2. Two terms t' and t denote the same entity if and only if a) their putative referents play the same causal role with respect to a network of phenomena; and b) the core causal description of t' takes up the kind-constitutive properties of the core causal description associated with t (Psillos 1999, 296).

It should be noticed that these preconditions on the worldly phenomenon of reference, and therefore truth, violate the Metaphysical Thesis as stated. Postulate (1) positively alludes to the phenomenon of description to explain the phenomenon of reference. This is not a mind-independent matter. Paradigmatically, descriptions we make and use to refer to entities are phenomena that involve mental intentionality, amongst other things.

It may be believed that Psillos conceded too much to the Descriptivist, according to which terms are defined by associated descriptions, and that this particular feature is not preserved if we adopt a pure version of the Causal Theory of reference. The Causal Theory of reference may be *prima facie* plausible as an attempt to do the explanatory job without the help of reference to minds. Pure Causal theorists try to do this by saying that all that matters in reference is a causal link between the thing referred to and the term used to refer, that can be traced back to when the object was named. But this is not the case. As the theory has been formulated, in particular in explaining what these causal connections between speakers come to, *intentions to co-refer* play a crucial role in the reference of terms (for example, Evans 1973).

The Causal Theory of reference has its modern roots in the work of Saul Kripke (1972) and Hilary Putnam (1975). The theory says that proper names and natural-kind terms work in the same fashion. There is a baptismal ceremony for an individual or natural-kind of thing, where they get their name. There may be some sort of description that gets associated with the entity through an association with the term that refers to it. However, the description is not analytically true of the entity, but is merely a reference-fixing device whose truth is contingent, unlike what Descriptivist theory says. As Psillos notes, this theory has the advantage of preserving reference-stability of terms when there are changes in theory. Tom may be able to refer to Bobby Sands after hearing something Al says about him, though he has never met Bobby Sands and all he believes about him is the false belief that Bobby Sands died in a car accident, for example This is made to come about by Tom’s ability to co-refer to the person to whom Al refers, and similar connections in term transmission exist between Al and other speakers reaching back to the beginning, when Bobby was given his name. In this respect, it is evident, the Causal Theory of reference presupposes elements of the mind, its abilities to intend, to explain the reference of terms.

However, the Causal Theory has several deficits by virtue of ignoring what the mind contributes in setting the reference of a term, if it construed as a non-mentalistic theory. Coming into causal contact with a thing named is hopelessly ambiguous, for unless one is dealing with merely a basic entity, in the impossible circumstance that would require, one is in causal contact with many other things and kinds of things apart from the thing one purportedly referred to in an act of baptism. Here I briefly elaborate various approaches to the point.

The Causal Theory fails to explain how it is that we refer to some things in the environment and not others by our term.[9] The Causal Theory of reference could just as easily assign “Aristotle” to one of Aristotle’s hairs when he was named, were it not for the intentions of whoever gave him his name, and it would then be deeply underdetermined when proper use of the term “Aristotle” is displayed and when not. Different interpretations of “Aristotle,” as of the man and of the hair, yield different truth-values to the assertions made using the term. The Causal Theory without further refinement just as easily assigns the same content to “having a heart” and “having a liver.” Further still, it fails to distinguish cases in which we introduce a new name for an entity from cases of mispronunciation. These difficulties can only be solved by reference to the psychological facts of the baptizer; for example what he or she intends,

[9] I go along with the simple story of term-reference as a function of a baptizing event, as does Psillos (1999, 297), amongst others; though of course, this is not the only relevant variable.

through her selective capacities, to give a name to, and the intentions of the speaker/thinker when s/he refers.

Further, ambiguity happens not only in the baptismal circumstance where a term is introduced but also in the circumstance of reference with a term already in use, where people in a linguistic community display their use of the syntactic item. Unless there are psychological facts about speakers mentioned in an account of their referential capacities with respect to a term of their use, it is deeply underdetermined whether they are deferring to experts, intending to co-refer with certain people at certain times but not at others, for example, or baptizing the thing they are predicating about with the introduction of a syntactically identical term. For example, “apple” is the name of certain fruits, but then it was also used to baptize the computer, the album, and perhaps even before the baptism of the fruits, to (sur)name certain portions of the population. These different uses will yield different truth-values to the assertions made using those terms. *Merely* syntactically identical terms cannot be substituted in assertions in a way that guarantee’s truth-value preservation—to suppose that the preservation of truth-value would be preserved in such semantical substitutions would, of course, just be equivocation. Causal Theories need psychological facts introduced into their theories to make the right distinctions about when we assert truths and falsehoods about our referents. This does not make them any less causal, it just notes that no viable theory of reference, the phenomenon which explains truth, will be without a mental contribution.

Further, it is desirable for a theory of reference and meaning to explain how we can be *wrong*. Such a theory would explain how it is that we apparently referred to things we eliminated from our ontology, and how it was rational to do so. The simple

Causal Theory has no say in this matter, since many terms that were introduced to later be eliminated were introduced in what may be idealized- as we allow to be done with other terms- into a baptismal ceremony for naming something that does in fact exist. And this would seem to commit us to not being able to eliminate many of the terms of our scientific discourse which we deem lacking of reference. So we would be committed to the existence of phlogiston, since "phlogiston" is the original name we gave to what we now call "Oxygen." What could make the difference between the reference of "Phlogiston" and that of "Oxygen"?

These are shortcomings that the descriptive element of Psillos's Causal-Descriptivist theory aims to make good on. According to the pure Descriptive theory of meaning, terms have descriptions associated with them analytically true of their referents. Objects are referred to by the terms just in case they satisfy the description, and assertions using the terms are true just in case the things they refer to make them true. Kripke and Putnam contended that these descriptions can be highly misguided and informationally incomplete, so we must not hold them to be true by definition. Psillos's theory combines the Causal and Descriptive theories into a theory that aims to make space for the right answers to the objections posed to each individually, expressed in the two postulates above. "Aristotle" refers to Aristotle and not his hair x, because "Aristotle" and "x" have different core causal descriptions associated with them, and the same goes for the "having a heart-having a liver" example, referents of "Apple" in baptizing versus deferential uses, and the case of "phlogiston" and "Oxygen."

According to Psillos, entities fall into a natural kind if and only if they have the same internal structure. The term that picks it out is associated with descriptions that aim to describe what this internal structure is like (its constitutive properties), and this can only be done by using the theoretical resources of the epistemic/theoretical environment in which the term is introduced and used. These ways of involving the mind in how we describe hold whether the external environment contributes in setting the meaning of the term in question as well. Thus, the mind is essential for reference, reference being a prerequisite for truth.

Now there is one other issue lingering: are the mind and its properties natural kinds of things? We must get clear about what must not be demanded of a natural kind in this context. If it is regarded as a mind-independent thing, then because psychology involves the mind, it is not mind-independent. Further, if certain psychological facts are determinative of the way psychological theory attains us epistemic access to the psychological reality under investigation, then by reasons rehearsed above, again, psychology is not mind-independent. This last reason is in the good company of physical truth or knowledge being mind-dependent. But to assert this is not idealism, it is merely to note a feature of the way truth and knowledge are constituted. Minds can also be regarded as natural kinds of things, as long as we understand natural kinds of things to just be things in nature- excluding the supernatural, as a contrast class. Such a conception would allow artifacts, like dams and rheostats, to be included as natural kinds. This would be unintuitive for many, since a paradigm contrast class of natural kinds is artifactual kinds. I suggest just to keep silent then on the issue of natural kinds- and just speak of nature. Like beaver's dams, our dams and the rest of our artifacts, can be included within it.

To conclude I propose that the following amendment to the Scientific Realist's first postulate:

1.' The Metaphysical Thesis: The things in nature, sometimes mental and sometimes not, we aim to refer to with our scientific theories, make our theories true or false.

It is noteworthy that idealists could not accept this thesis because it endorses the idea that there are non-mental portions of nature.

Its application to mind is:

1M. The Metaphysical Thesis about Mind: The mental things we aim to refer to with our psychological theories make those theories true or false.

1.2 Semantic Realism: Theories, Instruments and Visions

In this section I examine a particular way of understanding theoretical discourse. In particular, I want to argue that scientific discourse outstrips observational discourse in a way that has the ability to refer to unobservable portions of nature. I want to argue that reductive empiricist and instrumentalist interpretations of theoretical discourse do not have proper justification, while Semantic Realism does. In particular, I will use Paul Churchland's argument against Constructive Empiricism to analogically argue against the unjustified selective skepticism of the positivist understanding of theoretical terms. Irrealist alternatives unduly and unnecessarily complicate our semantic commitments. This analysis will also note the slippery slope derived from the criterion that something must be observable in order to be believed in, in the Constructive Empiricist case, and the criterion that truth-conditions must be about observables. The slippery slope leads to accepting only present sense-data as existents and statements truth-valuable only if about present sense-data. This extremely slim, and largely futile, conception of semantic content for scientific theories, or other statements, can be seen as a reduction to absurdity. Positivist interpretations of scientific language are undermined. However, the route to this conclusion does not wield the "has to be mind-independent" axe that Psillos wields for it, for as we have seen, such an axe is not only unnecessary but misplaced.

Reference to Mental Unobservable and Epistemic Inevitability

Mental subjects and properties are sometimes not directly observed. Normally we do observe people in the sense that we, for example, observe their faces. However, we do not observe people in that we do not directly observe their minds, and a person's mind's existing is essential for the person her/himself to exist.[10] We do not normally observe people's brains, neural-events, perceptions, beliefs, etc. Were someone to lose their face in an accident, they could remain a person and they could remain the particular person they were prior to the accident. The implication for psychological discourse is that it, like physics, has the capability to refer to those unobservable, or theoretical, portions of nature.

Psillos contends that positivist interpretations of theoretical assertions (t-assertions), assertions that have theoretical terms (t-terms) that seem, taken at face value, to typically make reference to unobservable entities, are not appropriate. As a better alternative, he advocates Semantic Realism, the view that taking the apparent reference to unobservable parts of nature that t-assertions and t-terms seem to have is the correct interpretation. Theoretical assertions are typically about unobservable entities standing in certain unobservable relations. The truth-conditions of t-assertions are (typically) satisfied when and only when unobservable reality is the way it explicitly says it is. By contrast, reductive empiricists say different things. Reductive empiricists say that t-assertions are translatable into observational assertions (o-assertions), assertions explicitly about the observable world. Their truth-conditions are satisfied when and only when the observable world is a certain way. Psillos accuses reductive empiricists of too much mind-dependence in their semantic criterion, since truth-conditions should be distinguished from verification-conditions. Tying truth-

[10] This is irrespective of whether one thinks that it is the Physical View of Personal Identity, where a person is taken to be identical with his or her brain, a view associated with Bernard Williams (1991). The assertion is irrespective of this position since the brain must be mentally functioning brain in order that we don't count the dead among the living.

conditions to verification-conditions seems to give *us* ontological priority in what the world contains, rendering it mind-dependent, he contends.

However, observable conditions do not require a mind. Were there no minds the observable conditions would remain the same: those things which would be observed if there were minds capable of external observation.

The reductive empiricist says that everything that is true is true in virtue of the putative fact that were we to verify it, we would observe the asserted objects and relations. According to reductive empiricists, the set of observable entities are taken to be epistemically privileged and we cannot even predicate anything about things that we do not stand in this privileged relation to (Psillos 1999, 13). This ties truth to our mental properties: our abilities to verify and observe. Psillos contends that which assertions are true are independent of this: "To say of a non-mental entity which features in one's ontology that it is *mind-independent* is to say that assertions about this entity are true because and insofar as their truth-conditions obtain, and not because and insofar as such assertion can be verified, rationally accepted, believed and cognate epistemic notions (Psillos 1999, 14). Psillos criticizes reductive empiricism because according to it, truth is partly epistemic.

Three criticisms of Psillo's line are in order here, it seems to me. First, at the very least in the case of psychology, it is misplaced to demand mind-independence, for as was discussed in the previous chapter, mind is the very subject of theorizing in this case. He is sometimes sensitive to an aspect of this fact, when he qualifies his above statement to be only about non-mental entities. However, what are we to say about the subject of psychological science? Since our abilities to verify, rationally accept, believe and cognate epistemic phenomena are themselves part of the subject of study here, the truth-conditions of statements that make reference to them obtain, or are made true, precisely insofar as our abilities to verify, rationally accept, and so on, determine and allow. It is trivially true that statements about our epistemic properties have their truth-values determined by our epistemic properties. Psillos does not seem to be able to recognize psychology as an objective science, were his principles generally upheld.

Look Again

Secondly, it is short-sighted to say that what matters are truth-conditions (that "the referred to entities stand in the referred to relations") *rather than* verification-conditions. For verification-conditions are just a sub-set of possible truth-conditions. Take the verification-condition to which a putatively "real" truth-condition might be contended to be reduced: *were we to look up 70 degrees at time t we would see a bright yellow ball we call "the Sun."* This is the reducing verification-condition of the putative truth-condition, say, *were we to look up 70 degrees at time t we would see a ball, one of whose layers undergoes the proton-proton chain emitting photons as a result, according to the equation* $E=mc^2$. It may well be that Psillo's Semantic Realist wants to say that the verification-condition specified above does not capture all the commitments of the second, supposedly exclusively real, truth-condition. But the fact that verification-conditions do not account for all the possible truth-conditions of scientific assertions (and reference-conditions of terms), does not imply that verification-conditions are not a variety of truth-conditions: conditions that can be true or false. For evidently, the above verification-condition can, as the referred to entities of this assertion would have to stand to the referred to relations for it to be true.

Independently of whether this verification-condition is the correct translation of the more theory-laden truth-condition specified after it, it nevertheless specifies a way the world must be in order for it to be true, the defining trait of truth-conditions.

Were we to say that verification-conditions are not a variety of truth-conditions we would have to say that all talk in observational terms not only is not, but cannot be, true or false, since talk in observational terms has verification-conditions as its meaning. This would just inflate theoretical discourse to the extent of displacing the more mundane observational discourse. Both kinds of discourses have their place, I think, and no argument has been given against the importance of either.

Don’t Forget About the Meaning-Makers Realism Requires

Thirdly, Psillos seems to think that truth-conditions are mind-independent (in contrast to verification-conditions). Truth-conditions, the sort of meaning at issue here, are things we give to our assertions with our mental and social capacities and practices.

Take for example the truth-condition of “electrons have a negative charge.” It is: *that electrons have a negative charge.* This is the interpretation we gave it, since there is nothing that prevents us from assigning to it the truth-condition *that electrons have nationalities.* This sort of assignment is a fact our mental capacities contribute, and is made quite explicit use of in the design of codes. Thus, whether the truth-conditions of our assertions and theories obtain, is not only partly a matter of the way the world is independent of the truth-conditions of our assertions and theories, but also partly dependent on what the character of the mentally determined conditions of truth pertaining to the relevant assertions and theories are. It follows from the fact that the truth-conditions required for scientific discourse are mentally determined, that meaningful theories, assertions, and terms are not radically non-epistemic, and therefore involve an epistemic component anyhow. That is, independently of whether reductive empiricism requires an epistemic or mental component, *Scientific Realism and particularly Semantic Realism already requires it.* So it cannot be a criticism of reductive empiricism, when Scientific Realism generally, and Semantic Realism particularly, are posed as the endorsable alternatives, that reductive empiricism ties scientific truth to mentally determined epistemic characters. So this argumentative strategy of Psillos's against reductive empiricism does not work.

Psychology=Behaviour, or Not

According to reductive empiricists psychology does exist. It is just that psychology turns out to be about something other than what we took it to be about. Particularly, it turns out be, not about the mind, but observable behavior.[11] All the t-assertions made in psychology are translatable to o-assertions, in this case, assertions about behaviour, reductive empiricists claim. The theoretical terms of psychology seem to make reference to minds and mental events minds take part in, having certain mental properties, and assertions using those terms are true because of the way the mind is. But the reductive empiricist claims that this is only appearance. The reality is that all that exists, pertaining to psychological discourse, is behaviour. The immediate objection to reductive empiricism is that t-assertions in *psychology causally explain* behaviour, and the reductive empiricist scheme undermines the explanatory edge of psychological ascription. The same can be said for t-terms in physics, which is what Psillos has most centrally in mind in the debate over Scientific Realism. Some theoretical entities of

[11] The prime example in the psychological variant of the reductive empiricist tradition is Carl G. Hempel (1977).

physics causally explains some features of observable phenomena, and to do away with reference to unobservables strips physics of being able to causally explain, for example, how stars are formed and why they shine.

Psillos emphasizes that reductive empiricists confuse truth-conditions with verification-conditions, resulting in a loss in explanatory power. What makes t-assertions true are things generally referred to in order to causally explain the truth-makers of o-assertions. If we equate t-assertions with o-assertions, there must be some explanatory motivation, which has not been given. Rather, it seems like we end up without causal explanations of the things referred to with o-assertions, in effect, losing explanatory power (though see below).

It should be noted that reductive empiricists can still say that the truth-makers of t-assertions are identical with the truth-makers of o-assertions since their truth-conditions are identical in the last analysis to their verification-conditions, and that for any *specific* explanatory context we can find appropriate o-translations. One bit of behavior can cause another bit of behavior, just like chairs can cause us to sit without falling through. This is all formulated in purely observational terms. Further, positivists, including reductive empiricists, were Humeans about causation, which just recognize that causation need not “make any sense” in that the cause logically necessitates its effect. Contemporary Humeans can assert that the simple constant conjunction conception is too blunt a conception to describe causal relations, but more complex functions construing constant conjunction properly do the job just fine. However, more must be demanded of this attempt at a come-back. The claim that in any specific context we can find a causal explanation of behavior by behavior via a function (whose additional terms are mathematical or logical rather than empirical) is false. Take the compelling idea that in some particular case a pain causes a wince. According to the behaviourist, this instance of pain is identical with this instance of a wince. So it is not the case that one causes the other, since identical instances cannot be related as cause and effect. This is bad news for the reductive empiricist strategy, for by analogous measures we can get the same result for physics, as well. If atoms don't cause the observational evidence we have for their existence, being identical with it, then it looks as though the reductive empiricist fails to causally explain the evidence Scientific Realists use theoretical commitments to explain.

Psychology=Mathematical Fiction, or Not

Instrumentalists had a similar approach to theoretical discourse. According to instrumentalists theoretical terms are calculational devices, which are useful in predicting observation, but do not refer to entities. Which function is assigned to each term is determined by the usefulness of the function in successful prediction, but in so doing we are not ontologically committed to the existence of an entity to which the t-term expressing the function corresponds. In psychology, the implication of this construal of psychological theory is that its apparent reference to things like beliefs, perceptions, minds, and so forth are referentially empty and the apparent assertions employed in scientific contexts do not have a truth-value. Terms and assertions that are employed to describe the causal mental origins of a bit of behavior are not really describing anything in the world. They are mere pieces of syntax, without referential capacity, we use to predict observable behavior from behavior.

In fact, instrumental behaviorists and their reductive empiricist peers can meet the explanatory challenge, and have an answer to the above example of the pain and wince. They can say that a particular pain does explain the wincing. Their theory does this by saying that a crucial part of the pain-story above is incomplete: the destructive impingement of some object on the skin for example. Instrumentalists can say that the function, <pain>, takes in descriptions of external causes of pain and outputs descriptions of winces. The reductive empiricists can say that “pain” just specifies ordered pairs of destructive impingements and winces, and that the wincing is explained by the subject’s exemplifying of the ordered pair characteristic of “pain.” Let us suppose that such an empirically adequate function and specification could be invented.[12] In that case the positivist makes good his claim that he can predict and so explain, via this analysis, what the realist about pain can, without ontological commitment to unobservable pains.

The problem for this proposal is that it makes an unwarranted move. It tells us that skin-impingements and winces have reference, but pains do not, but why this is so is not explained. “Skin” and “destructive impingement” are terms whose correct use is theoretically guided by scientific exploration. At least, they are *not more* observational and less theoretical, than “big radioactive rock.” If the distinction between “pains” and “skin” is attempted to be justified by the idea that skin is observed, we can counter that we can perceive (observe) pains at least to the same degree, either by seeing someone in pain, just as we see people with skin, and by feeling it ourselves. The case is analogous to the observation of big radioactive rocks. Such rocks are observable, but

[12] “Empirical adequacy” is a term of van Fraassen’s, used to mean that it implies true observational descriptions.

have unobservable nuclear structures that make them radioactive, and which can be cited in the causal explanation of observable features of the rock.

Are We Stuck in Wittgenstein's Box?

Some may have doubts that we can recognize the private states of mind of others and that we can unambiguously be said to be in a state of the same kind when we say that you and I are in the same private state of mind, such as being in pain. The positivists might have had their doubts themselves guiding their theories, since their guru, Wittgenstein, expressed doubts of this kind. Wittgenstein had a colourful fable about why certain psychological phenomena are not referred to in a successfully communicable way and therefore we cannot talk meaningfully about it.[13] Suppose that you and I have a box in which we each have an object. You and only you can look inside your box, and I and only I can look inside mine. The objects in the boxes we will call "beetles." It is quite possible that when trying to communicate with each other in saying "I have one beetle in my box" we are referring respectively to a mouse and a lizard. Our purportedly internal mental phenomena, like having a pain, is like your having the mouse and my having the lizard, in that they are private in a way that is only accessible to each of us individually. To say anything about them in a public language is bound to miscommunicate, Wittgenstein seems to argue. Since terms that cannot communicate, like "beetles" in this context, "pains" in psychology, along with other particular psychological terms, do not have factual reference and sentences using these terms are therefore not truth-valuable, unless they are provided with an interpretation

[13] As depicted in Carl G. Hempel (1980). The reader may note the apparent inconsistency in saying that one cannot talk meaningfully about certain mental states, but thereby, in saying so, doing so. I am told that Wittgenstein was okay with saying he was not speaking meaningfully. Some of us do not find it unintuitive that in saying ""gibberishly-doo-dah" is meaningless," we speak meaningfully.

that does away with their incommunicable quality. This is what the reductive empiricist and instrumentalist semantics can seem to provide.[14]

Here's the Way Out

We might question, however, whether we are actually *bound* to miscommunicate about what we begin to call "beetles," and therefore whether the analogical force of the fable is lost. For example, we could communicate and investigate into whether we are talking about the same sort of thing, by exchanging information about whether our corresponding beetles move, whether each is a biological creature, a mammal, four-legged, furry, what we feed them, and so forth. By a series of questions and answers you and I could discern what each has in each's box and the ambiguous way in which "beetle" was introduced would be disambiguated. Surely we could try to deceive one another, but there is no *a priori* reason to think we are bound to fail to find what the unobservable object the other calls "beetle" is, unless we postulate severe constraints on theorizing and experimentation that we don't have *a priori* reason to think we are constrained by in the science of the mind. Surely, the process of investigation "without looking" is not 100% epistemically reliable- no process of investigation is. But there is no special reason to believe good reliability is not actual or attainable. In fact, we have reason to believe that such reliability is actual or attainable given our similar evolutionary background, biological constitution, successful communication about all sorts of things we attain public access to, and observable confirmation for.

[14] The 'beetle' literature is large. I am not addressing the whole of the Wittgensteinian problematic here, but only using it to make a point about realism in psychology in accordance with one possible interpretation of Wittgenstein.

Empirically Equivalent Contenders

The most prominent way in which positivists (reductive empiricists and instrumentalists) have tried to justify their differential semantic treatment of observational and theoretical terms is by allusion to the possibility of seemingly alternative theories that explain the same observable phenomena. If we take all such alternative theories to be semantically different, positivists believe we are inevitably left without a decision as to which of the empirically adequate theories to adopt. Their solution, whether of the reductive empiricist or instrumentalist sort, to this problem is to say that these apparently distinct theories were not really descriptively distinct (Sklar 1992). A prominent way to support the positivist interpretation of theoretical terms and assertions is by reference to Craig's Theorem. Craig's Theorem proves that given

> any first-order theory T and given any effectively specified sub-vocabulary O of T, one can construct another theory T' whose theorems are exactly those theorems of T which contain no constants other than those in the sub-vocabulary, O. Hempel was the first to recognize the broader significance of this theorem: for any scientific theory T, T is replaceable by another (axiomatisable) theory, Craig (T), consisting of all and only the theorems of T which are formulated in terms of the observational vocabulary, Vo (Psillos 1999, 23).

Hempel presented the significance of the theorem in terms of a dilemma, the *dilemma for the theoretician*. This dilemma says that either theoretical elements of theory either serve the purpose of deductive systematization of empirical consequences or they don't; either way theoretical elements are dispensable.

> (I)f the theoretical terms and the general principles of a theory do not serve their purpose of a deductive systematization of the empirical consequences of a theory, then they are surely unnecessary. But, given Craig's theorem, even if they serve their purpose, they can be dispensed with since any chain of laws and interpretative statements establishing such a connection should then be

> replaceable by a law which directly links observational antecedents to observational consequents.[15]

This consequence boosts the positivist cause since it seems like theoretical assertions and theoretical terms are dispensable. So suppose that there is an effectively specifiable psychological theory, T. Then there is a distinct, Craig T, which eliminates all the terms that make apparent reference to minds and mental properties and events, which has all the observational consequences of T. Therefore, by eliminating the psychological theoretical terms (either by a reductive empiricist or instrumentalist method), we "save the phenomena" of psychology without minds, mental properties nor mental events.

Deciding on Theory Based on Unifying Power: A Form of Evidence

However, Psillos argues that the Craigian strategy for dispensing with t-terms fails because Craig transforms lose the inductive fertility of their originals, and therefore that a decision can be made as to which empirically equivalent theory, T or Craig T for example, to choose (Psillos 1999, 25-26). It is customary and fruitful in scientific practice to take two theories T and T' (which share vocabulary), and get novel predictions that each one did not entail alone, and it is unlikely for this to happen with their Craig transforms. Theories imply their observational consequences but not the other way around and therefore successful results of theory combination are unlikely to happen if the combination is different. Thus, candidate Craig T joined with candidate Craig T' are not likely to give rise to novel successful prediction.

[15] Psillos (1999, 23) quotes Carl Hempel (1958). We can also work against the view assumed here about how predictions are gathered from theory, which is better expressed by Quine's "total sciences," but for the purposes of this essay I assume the positivist assumption that empirical consequences are mere "deductions" from theory.

If we regard T's and T''s contents to be exhausted by their *individual* observational consequences then we cannot expect to get the novel predictions we get when we combine them. But we do get such novel predictions! The theoretical part of theory is combined to establish inductive connections between the domains of the theory, and generates test implications the theories did not generate individually. There is no reason to think Craig transforms of T and T' will generate the sort of inductive integration and fertility of theory of T and T' combination (Psillos 1999, 26). In short, T and T' have much more *unifying power,* than Craig T and Craig T', since they are poised to become theoretically integrated, yield new predictions, and give rise to test-implications which if confirmed warrant them more than their Craigian alternatives, which do not receive these meta-theoretic boosts characteristic of these kinds of novel prediction.

It is hard to say why the observational domain is any more real than the domain of unobservables, such that only o-terms have reference and t-terms may have reference only insofar as they refer to things in the observable domain. We make fruitful explanatory use of the theoretical vocabulary to causally explain phenomena in a way that supports the making of inductive connections between theories, which has the consequence of unifying our picture of the world. This is something the instrumentalist is unable to justifiably do or explain.

The Realist Opiate- Or is it Just Oxygen?

However, it is open to the positivist who wants to base his Irrealist Semantic thesis on the basis of the fact that alternative Craig theories are empirically adequate, to say that for purposes of theoretically capturing more theories which have a wider empirically adequate scope, we must hold our theories (t-terms and t-assertions included) to be tentatively true. In this way, they serve the role of theory expansion that is required for scientific progress. However, at the end of inquiry, we no longer have that motivation. At the end of inquiry, there is no progress in fact to be made so it is simply not the case that we need to hold onto the theoretical aspects of our theories at the end. At that point, because we no longer find our theoretical ladder truly useful, because of reasons of the Craig alternatives, we can let go of our ladder and throw it away. Philosophers might look down in the following manner "Let the Philosopher rejoice on the backs of current science, raising only observational assertions to the throne of Belief, while the working scientist smokes the opiate of theoretical realism that moves him to get us Both to the Top through the Ladder of Deceit."

Let us evade the very concrete question about whether we will in fact reach this final scientific state, for a moment, through the series of scientific historical transformations. We must note that this positivist strategy just is Bas van Fraassen's argument for Constructive Empiricism. Instead of arguing for an interpretation of theoretical discourse, Constructive Empiricism is about whether to believe the truth of scientific theory. He argues that only belief in the truth of the observational parts of scientific theory is justified. Because there are always alternative theories with identically true observational parts, van Fraassen think we should not believe in the theoretical parts. That is to say, van Fraassen uses the same premises to justify a conclusion about what is believable to be true in scientific theory, while the positivists use them to justify a conclusion about what truth-conditions of scientific theory may be entertained in the first place.

My strategy is to show why van Fraassen's anti-realist epistemology is wrong and say that the same reasons show that the irrealist positivist semantics is wrong. Van Fraassen believes that the ultimate purpose of science is to give us an empirically adequate theory of the world- a theory that saves the phenomena, in his sense, a theory which gives us the right observational results (and nothing more, i.e., theoretical claims are not worthy of belief). His argument is just that which motivates positivist semantics, through Craig's theorem, namely, that if there are non-identical theories with the same observational results, it is arbitrary to choose any. From this van Fraassen draws the epistemic conclusion that we ought to *only* be ontologically committed to the truth of the observational portions of theory, while positivists conclude that only o-terms and o-assertions have referential meaning (t-terms and t-assertions have it only insofar as they are translatable into o-language). Both the positivists and van Fraassen hold that theoretical portions of scientific theories do not commit us to a theoretical ontology. The reductive empiricist does this by construing the apparently distinct theoretical assertions as ultimately saying the same thing about observational phenomena in different languages, while the instrumentalist does this by saying that apparently distinct theories with identical observational results are only truth-valuable with respect to their observational assertions, and therefore have the same truth-conditions, and therefore mean the same thing.[16] Van Fraassen does this by fiat.

Why the Arbitrary Selective Skepticism? The Case of the Future

An overarching distinction in the debate over Scientific Realism is that between parts of scientific theory which are properly ontologically committing and parts of

[16] We should note also that Craigian alternatives will always have an infinite number of axioms and so are in this respect, infinitely more complex than our theories, given that the latter are premised on finite numbers of axioms.

theory which are not. Whilst realists hold that the terms in a theory are all ontologically committing, irrealists like reductive empiricists, instrumentalists, and Constructive Empiricists, claim that only those terms about observable things are. Commonly, the relevant micro world is held to be theoretical and not observable, and Constructive Empiricism prescribes that in adopting a scientific theory, we do not commit ourselves to the unobservable entities of the theory. Paul Churchland (1982) argues that van Fraassen’s theory employs selective unjustified skepticism. He explains that the assertions about the future are theoretical as the future is not-now observed. There are many alternative possible ways the future might be and our assumption that it will be a certain way is a theoretical inference of the same type that is involved in making an inference about other theoretical entities. Hume makes poignant the fact that such inferences about the future are theoretical with his Problem of Induction. So we have no reason to doubt the existence of theoretical entities, like micro-entities, whose existence we infer without also doubting the existence of future entities, even the commonsense dry goods of tomorrow. To be skeptical about one kind of theoretical entity and not the other is *ad hoc*. In other words, Constructive Empiricism is unjustifiably selective in its skeptical scope.

Why the Arbitrary Selective Skepticism? The Case of the Past

However, there are also past events posited by theory that are not relevantly observable and are therefore theoretical, as we relevantly cannot go back in time to observe them, like the past existence of dinosaurs and even the activities we carried out yesterday. These are also unobservable from the point of view of anyone who might wonder whether dinosaurs existed and whether they met a given person the day before.

There are alternative ways the past might have been and to those ways there correspond alternative theories positing different entities, events, etc. There could have been Adam and Eve instead of dinosaurs and it could have been the case that we did not meet the person we in fact remember meeting. By the reasoning that van Fraassen employs, we it is correct to remain agnostic about whether dinosaurs existed and whether we met a given person, independent of the strength of the available evidence.

Van Fraassen: For Humans, Pestles are Unbreakable

Churchland's argument against Constructive Empiricism and my present extension of it might be questioned. For van Fraassen thinks that "unobservable" is not "unobservable in principle" as Grover Maxwell (1962) recommends, and then uses this conception to argue that since no theory implies that its entities are unobservable in principle and that increasingly better measuring devices expand our observational capabilities, nothing is unobservable. Rather, van Fraassen argues that "unobservable" is what is not observable by humans *qua* humans, that is, what does not come under the purview of our description by a final physics or biology. He says to Maxwell:

> This strikes me as a trick, a change in discussion. I have a mortar and pestle made of copper and weighing about a kilo. Should I call it breakable because a giant could break it? Should I call the Empire State Building portable? Is there no distinction between a portable and a console record player? The human organism is, from the point of view of physics, a certain kind of measuring apparatus. As such it has certain limitations- which will be described in the final physics and biology. It is these limitations to which "able" in "observable" refers- our limitations, *qua* human beings (van Fraassen 1980, 17).

Van Fraassen may try to argue that the middle-sized, dry goods of tomorrow are observable to us *qua* humans, whereas putative entities like electrons are not.

Ricardo: Yes They Are

Two points are in order: First, it seems to me that van Fraassen changes the subject matter. The truth of whether a mortar and pestle made of copper and weighing about a kilo is breakable depends on the purposes and means, including those created by humans, inherent in the context at hand. Let us grant van Fraassen the uncontroversial presumption that neither he nor I could break the pestle. *Qua* human they are unbreakable in this very constrained way. But by the use of a bomb already present in the war industry it is. Suppose that breaking the considered pestle has benefits for a military already having such a bomb available to it – the benefits, say, are saving hostages, obtaining a reputation for high defense capability, etc.- that outweigh the costs- the bomb, the transportation, etc. It follows that the officer, convinced by van Fraassen, that informs his superior that the considered pestle is unbreakable, is wrong. In this context, since it is humans and their artifacts doing the work, *qua* human, it is breakable. The question here is whether in the present context the question is relevantly truly answered by being interpreted more like the first case than the second. But it is clear that it is more like the second, lest we then, by the same standard, say that we cannot go at speeds higher than 25 kms/hour from A to B, that we cannot fly to Belfast, all in all, that we cannot do the typical things we increasingly are able to do with the help of the inventions we humans designed for such purposes. Perhaps it is not news to anyone but to van Fraassen that console players are portable, though they are more burdensome to carry can others. In ordinary contexts, the less burdensome to carry ones are called “portable,” while the more burdensome, “not portable.”

The modal force of uses of these duals is evidently weak in this pedestrian context, but it is also not the only context and clearly the debate over Scientific Realism demands of stronger modal interpretations. This is just the kind of stronger interpretations that allows for the expansion of observables with the aids we design for this purpose, just as we allowed when we consider whether a pestle is breakable in the mentioned military context.

Secondly, humans are social animals and have histories of their science and technology. It is baseless to exclude these human properties from the purview of what we can do *qua* human, while including physical and biological properties. That our powers of observation have been expanded, just as our powers of breaking things, is easily shown by how our microscopes, telescopes, and the rest of the plethora of scientific measuring devices available to us today do just that. If from the point of view of physics, we are a kind of measuring apparatus, as van Fraassen contends, and our observations are acts of measurement, then like any other measuring apparatus, its measuring capability can be expanded and contracted, and other components may be added for optimizing measurement of physical properties.

If the reader has a problem with comparing what is observable to what is breakable, remember that it was van Fraassen who introduced the comparison to make his case. All I do is show that even if we do not challenge the adequacy of the comparison, it still falls prey to the realist assertion that observation and measurement expand.

Van Fraassen can also argue that we can observe the future by “waiting for it,” and that therefore the entities inhabiting the future are not theoretical. However, the debate over realism should tell us about the kinds of things we are committed to in virtue of our commitment to a theory at a particular moment, since commitment to a theory at a particular moment is the condition in which we find ourselves when considering such things.

For van Fraassen, it is even harder to justify the idea that the entities of the past are not theoretical. The idea that we can wait to observe the entities of the future does not even arise in a supposed attempt to justify why to be committed to the entities and events of the past. Perhaps we can time-travel to the past, and as we shall see I will argue that we have reason to believe we can, but this capacity is way out of our present reach: way more than flying to Belfast or observing a neuron. We don't have the technology and much less can just expect to be able to wait for it as we are able to wait for tomorrow. There are many ways the past could have been, compatible with our present intrinsic observational state, and to these ways there correspond alternative theories. If van Fraassen's reasoning is correct, then we should not be committed to any truth about the past, even by his quasi-coherent standard.

We Don't Observe Theoretical Entities, We Just Observe Their Effects, Or Not

Van Fraassen or some other irrealist may argue that we do not observe neurons *but their effects.* While this is true, it is just a property of observation more generally that we observe external things by observing their effects. This is seen by considering any putative middle-sized dry good that is paradigmatically an observable object: a tree, say. Suppose we saw it. Well we would have the same perception, *as of a tree,* in the alternative scenarios where it is clear that it is effects that that are sustaining perceptions of it. If the tree disappeared but the neural processes in the brain of the observer are maintained, say by spontaneous dream-like activity, or by replacing the tree by a holographic illusion of it, the perceptual content, *as of a tree,* would be maintained. (For externalists who might deny this, it is a challenge for them to account for false perception.) Consider also that were the tree observed at t, there is a scenario in which it is spontaneously obliterated and there would be a t+1 such that the perception *as of a tree* is present and the tree is not (given that (t+1) - t is the time it takes for photons from the tree to hit our retina). Consider also the observation of stars that are dead and no longer producing light. In those cases we have perception of stars, without the stars themselves. It is their effects that maintain the perception. In normal cases of perceiving trees, or anything else, the lag is just not so apparent, but is nevertheless real for it takes some time for the information to get from the object of perception and the organs of the perceiver. We observe things by observing their effects, so to deny observation of theoretical entities, like neurons, atoms, or any other theoretical entity, because all we observe is their effects is a red herring.

Exclusive Commitment to the Specious Present? Not Me

Notice that this argument is effective against the possible challenge posed above, that even if commitment to entities posited by theoretical aspects of scientific theory is necessary for the development of science, we need not be committed at the final epistemic stage of scientific development, as long as we kept on living. Suppose we arrived at a good adequate theory of everything with no necessary modifications for a year. Then, van Fraassen would restrict belief to entities that can be observed arguing that the evidence is compatible with many theories. Paul Churchland points out that there is selective skepticism involved in holding that the observable predictions about the future will turn out to be true, while other unobserved aspects of nature posited by the theory are not (the parts about micro-physics for example). For the future also, we find radically different theories whose predictions about the future are correspondingly radically different, but van Fraassen is not prepared to be skeptical about the future with respect to the observational aspects of T. These alternative ways things might go is just the basis of Hume's Problem of Induction, which, tough as it is, does not warrant the conclusion that we should not form beliefs about the future or that we have no knowledge of it (though Hume may disagree!). However, by van Fraassen's reasoning, we ought not form beliefs about the future either even in this case where we have a putatively final empirically adequate theory available.

Further, van Fraassen cannot hold with any justification that the final theory in question is empirically adequate with respect to the past either, since this also would involve theoretical commitments about the trustworthiness of memory and testimony in judging T to be empirically adequate with respect to the past. If we take a radically different theory to evaluate the character of our world of the past, the apparent trustworthiness of actual memories and testimonies of the present are equally, radically possibly false. We can imagine Russell's scenario of having been created just five minutes ago to correspond to one such possible theory, according to which the past is radically different from what we believe and the present is intrinsically the same. Thus, if we employ van Fraassen's line of reasoning consistently, we see that all we are left with as the result is that the exclusive object of proper belief is the "specious present."

The line of reasoning adopted by van Fraassen leads to the reduction of what we take to be reality, by Cartesian alternative hypotheses, to present sense-data. For clearly, as Descartes showed, the existence of vivid dreams provide a perfect alternative hypothesis for the explanation of present observation. A theory that posits dreams as the stuff of observation is compatible with what is observed, therefore, by the reasoning advocated, it warrants for van Fraassen, the conclusion that we remain agnostic about whether we are in a dream and hold belief only about present sense-data. This is clearly too little to settle for, and we are led here by the principle that if we can find an alternative theory that explains “the data” don’t commit to anything but the parts of the theory describing “the data.” But since van Fraassen wants to believe in portions of reality expressed by the theory in question that go beyond current sense-data to allow for belief in what the theory says about the future, past, the existence of external middle-sized objects, and other minds (one would think!), to then withhold judgment on aspects of theory and world because of alternative hypotheses, is just *ad hoc*. This is a second extension of Churchland’s argument against Constructive Empiricism.

The Trilemma for the Anti-Theoretician

The criticism then can be cast as a trilemma, the Trilemma for the Anti-Theoretician: Either the ontological commitments had by virtue of the acceptance of scientific theories apply across micro, past, future, and other theoretical domains, on a par with the observational domain, or the Anti-Theoretician is utterly *ad hoc* in withholding judgments about his or her theoretical commitments, or is properly said to propose to be committed to nothing in holding theories, for example Einstein’s theory of relativity, but current observations. All the anti-theoretician’s options seem a little too implausible (the first one is implausible only if he is to remain an anti-theoretician).

Now Churchland’s argument and my two extensions have definite analogues in arguing against positivist irrealist semantics. Here what is at issue is not the truth of theory, but the truth-valuability of theory. Take the positivist principle that if there are putative theories differing in their assertions about unobservables, compatible with the same set of observation-sentences, then the theories’ truth-conditions are identical. However, alternative sentences about the future are always available, compatible with sentences about current observation, which is what we always find ourselves in a position to observe. Given these alternative sentences about the future, by positivist standards, sentences appearing to assert something about the entities in the future do not manage to assert (without reduction to observation-sentences about the present). By extension, given that there are alternative incompatible sentences appearing to be about alternative ways the past could be, these are also not really assertible (without reduction to observation-sentences about the present). By extension again, given that there are alternative incompatible sentences appearing to be about alternative ways the external world, including other minds, could be, these sentences are not really assertible (without reduction to observation-sentences about the present).

I conclude that the positivist has not properly motivated theoretical assertions to be dispensable in order to motivate his semantic thesis, and has not given adequate reasons justify the move from, “alternative theories exist,” to “we are not justified in believing one rather than another.”

I propose the following reform to Psillos’s version:

2.’ The Semantic Thesis: Scientific theories are truth-conditioned descriptions of nature, including portions of nature that are not objects of observation.

2M. The Semantic Thesis about Mind: Scientific theories of mind are truth-conditioned descriptions of mental portions of nature, including mental portions of nature that are not objects of observation.

1.3 The Epistemic Thesis: Computational Cognitive Science

The Epistemic Thesis is the hardest component of Scientific Realism to defend, and there are a plethora of dimensions that might be pursued to evaluate its truth in application to psychology. It is, nevertheless, a central component of Scientific Realism, and Mental Realism, as one of its domains of application, must say something on this issue. It is hard to defend because we do not want to be committed to the complete truth of our conceptions of mind at the moment.

Thus, in the main, I will concentrate on an argument that arises in the philosophy of mind against the truth of a particular scientific conception of the mind: that pertaining to computational cognitive science. However, another dimension of the truth of theoretical statements about the mind has already been addressed in the previous chapter. If we acknowledge that psychological theory has truth-conditions, and we accept that there is good reason to believe in the existence of the unobservable portions of nature that satisfy some of those truth-conditions, and in the observable portions of nature that verify theoretical truth-conditions, there is no good reason to exclude psychological theoretical terms from the ability to refer nor to exclude psychological theoretical assertions from being true.

1.3.1 Is Computational Cognitive Science a Field Born Dead?

John Searle is a scientific realist, and encourages cognitive scientists to be “interested in the fact of internal mental states, not in the external appearance,” their “obsession with method, and the tacit rejection of truth as the aim of investigation” has engendered. “Our claims, if true, have to correspond to facts in the world” (Searle 2002, 61). As a scientific realist Searle must believe that the scientific method is a particularly good tool for getting at the facts, i.e. that the scientific method reliably delivers claims that "correspond to facts in the world." The properties such deliveries posit are objective properties of the things that have them. However, cognitive science has gone astray, he argues, because it has embarked on a path that characterizes mentality in a way that doesn't correspond to reality in an appropriate way. Computational cognitive scientific theory, if it is to be about the mind, must correspond to mental reality generally, and mental content particularly. But according to Searle, computational approaches to mind, just cannot, logically speaking, do that. Computational theories of mind are born dead, in that we know from the beginning, that they will not correspond to the mental facts.

There has been a great amount of discussion of Searle's arguments against computational cognitive science. The main purpose here is to construct an argument based on Searle's work, though not completely explicitly formulated in it. It is an argument of which Searle endorses the premises and conclusion, but has not formulated in sequential manner. I hope the exercise will acquaint us with a not-so-explored dimension of Searle's problem with computational cognitive science. There are a few fill-in-the-blanks that are left for us to reconstruct. I hope the way I fill-in-the-blanks in the narrative constructed from his remarks is not controversial. I will subsequently go on to suggest how his worries are overcome.

But first, a few remarks on the debate Searle has spurred.

1.3.2 The Kind of Argument Searle Employs

Searle sees his hammering at the foundations of computational cognitive science as a logical rather than empirical exercise to knock-down the assumptions grounding a scientific field gone astray, as Scientific Realism allows. It has gone astray by just ignoring self-evident or logical truths about mind and computation.

> When I say that the implemented program by itself is not enough to constitute consciousness and intentionality, that is a logical claim on my part. By definition, the syntax of the program is not constitutive of the semantics of actual thoughts (2002, 56-57).

Epistemically, the claims of computational cognitive science are *a priori* known to not correspond to the facts. Logically, having computational properties of mind does not imply having certain mental properties. Metaphysically, having computational properties is not sufficient for having mental properties.

Searle's well-known strategy is to make a logical point about what it is *not* to have mental properties, based on a particular conception of what it *is* to have mental properties. More specifically, he claims that to have mental properties is not to have certain computational properties, as mental properties have a semantic content, the mark of the mental as Brentano had it, computational properties just don't have. The thesis that he aims to refute is Strong Artificial Intelligence (AI): "thinking is merely symbol manipulation…The mind is to the brain as the program is to the hardware" (Searle 1990, 116). A few remarks are in order.

First, Searle has no problem with saying that the mind is a computer, because he believes that everything is a computer (and any computer), but believes that this is no non-trivial truth. The problem is that being any possible computer is not logically sufficient for being a mind, and implementing any computation is logically insufficient for having mental properties. There will always be a logical/ontological gap between computing and cognizing, he believes.

Secondly, it is tempting to interpret Searle's characterization of computational cognitive science as asserting that for every mental property there is a computational property to which the mental property is identical. But perhaps this reductive identity claim is indeed quite strong, and therefore has less of a chance of being true. All that computational cognitive science needs is a computational sufficiency thesis, while being silent on type or property-as-such identity.

The Computational Sufficiency Thesis is such a slightly weaker thesis, and it says that there is class of computations such that for any machine implementing them has the mental properties at issue. David Chalmers formulates the Computational Sufficiency Thesis as the assertion "that there is a class of automata such that any implementation of an automaton in that class will have the mental property in question" (Chalmers 1996a, 309). Automata here should be understood as computers- entities whose fundamental feature is to compute certain functions. It is this thesis that is the crux of disagreement between Searle's position and computational cognitive science.

In the present context, there would have to be a class of properties here normally catalogued as mental that would have to be classed as irrelevant if the thesis is to hold. Notably, those mental properties whose obtaining depends on their being facts extrinsic to the thinker would have to be excluded. For example, knowledge that some external property is exemplified would have to be excluded since external knowledge requires that the external property known be indeed externally exemplified and computationally equivalent systems can vary in this regard. Also, certain indexically given knowledge might have to be excluded. For example, the thought *I am in Ireland* could be equivalent in terms of its computational properties but the value of "I" would vary depending on who is thinking the thought and thus there is a way in which different thinkers think different thoughts in thinking *I am in Ireland*. I ignore this kind objection to the Computational Sufficiency Thesis. It is a problem, I think, not particular to Computational Cognitive Science, but quite general, for any theory that is particularly of mind. Further, some computationally equivalent systems might not be psychologically equivalent in that one system in question is too slow to survive or pass an IQ test even though it computes the same functions as one that does. These cases show that the Computational Sufficiency Thesis needs refinement. The sense at issue here is the sense in which two systems are said to be psychologically equivalent if they share the internal mental features of their mental properties. This is the sense Searle wishes to consider.

1.3.3 The Reconstruction

One may extract from Searle's work the following argument against computational cognitive science.

1. Mental Realism: Mental properties are objective properties of entities.
2. Computational Subjectivism: Computational properties are not objective properties of entities.
3. Non-Objective Insufficiency: An entity's implementing non-objective properties is not sufficient for the implementation of objective properties.
4. Therefore, an entity's implementing of a computational property is insufficient for the implementation of a mental property.

The argument is valid. If you deny the conclusion and accept the premises, you get a contradiction. For then there would be non-objective properties, e.g. computational properties, the implementation of which would be sufficient for objective properties, such as mental properties.

A Threat to Mental Realism

The argument is of interest to those concerned with the reality of the mental, i.e. those who agree with Searle in asserting Mental Realism. The reason is that a way to deny the conclusion is to deny the Mental Realism that grounds it. Someone who takes this option would say that computational cognitive science is the best relevant theory we've got and that it entails that mental properties are not objective and therefore there is no barrier to computational properties, being non-objective, to be sufficient for non-objective mental properties. This option has the obvious disadvantages, to use Searle's criteria for being objective (more on these criteria later), that computational cognitive science would not posit intrinsic, empirically discoverable, nor causal properties. Thus, there would be no objective fact of the matter as to whether someone has a mental property in question, there would be no objective fact of the matter to discover through our best methodological practices, and minds would not make a causal difference in the world. What is wrong with the argument then? To guard against this threat to Mental Realism my strategy is to undermine the grounds Searle gives for believing in Computational Subjectivism.

Searle's Mental Realism and "Objective Properties"

Let us first see that the theses are found in Searle's work though. There are several ways Searle expresses his Mental Realism and from it we can learn about what he uses as criteria for being objective. He says, "there is no question that machines can think, because human and animal brains are exactly such machines" (Searle 2002, 56). Searle thinks an essential property of the human mind is that it has mental contents, intrinsic intentionality, which he expresses with the epistemic force of an axiom. He believes that any attempt to discover the nature of (at least the human) mind will have to discover the nature of the semantic features of mind, intrinsic intentionality, and any replication of a mind will have such features (Searle 1990, 117).

Mental properties of machines are conceived of as physical, on a par with the objectivity of electromagnetic magnitudes, molecular properties, lactation, digestion, and the weather (Searle 1990; 1993; 2002).

We can identify several conditions on something being objective in the way the mental is, that Searle countenances:

1. That mental contents are *intrinsic properties* of the organisms that have them.
2. That mental properties are empirically *discoverable*.
3. That mental properties are *causal*.

For mental properties to be objective they must be the way 1, 2, and 3 says they are. Mental properties are said to be objective when these conditions are satisfied. There is no unequivocal evidence as to whether Searle thinks, perhaps tacitly, that failing in satisfying any of the criteria for objectivity, while perhaps not failing in all, is sufficient for not being objective. Searle does think that computation fails each.

A quick note: I have qualified “being objective” to be “being objective in the way the mental is.” This is to guard against the idea that the mental is objective in the way in which pure mathematical properties are objective. This is because we definitely think that mental properties of entities in the physical world, the world in which we find ourselves, are not independent of how that physical world is. On the other hand, pure mathematical properties are the way they are independent of how the physical world is. That is, for example, whether such and such a notional Turing machine executes a given algorithm is independent of how the physical world is. This is analogous to the case of geometry. Interpreted as a pure mathematics, Euclid's axioms are true. Interpreted as the fundamental structure of physical space, they are false. But the fact that Euclid's axioms as about physical space were refuted, does not entail that they are refuted for mathematical space as well. For the question of whether the psychological properties of the brain are computational it is in this applied sense that is the relevant interpretation.

Searle holds that Computational Subjectivism is true. Searle aims to undermine the objectivity of computational properties of machines by saying that they do not satisfy each of 1, 2, and 3. There are four prominent ways Searle alleges computational properties fail to satisfy criteria for objectivity. One is that computational properties have only non-objective intentional properties because they are derived from us, and therefore are not intrinsic, discoverable, nor causal. Second is that computational properties are not objective because they are purely formal, and therefore not intrinsic, empirically discoverable, nor causal. Third is that computational properties are not objective because they are multiply realizable, and therefore not intrinsic, empirically discoverable, nor causal. And fourth is that computational properties are not objective because they are at best simulations of objective properties, and therefore are not intrinsic, empirically discoverable, nor causal. Next, we establish the grounds for attributing these charges to Searle.

Searle: Computational Properties Have Only Derived Intentionality, and Therefore Not Objectively

Let us begin with the first charge he makes, that computational properties have only non-objective intentional properties because computational properties are derived from us, and therefore are not intrinsic, empirically discoverable, or causal. He says that the intentional properties of computations are at best like the intentional properties of natural language utterances, such as the intentional properties of "es regent" in his example below. He contends that such a property of this bit, like any other bit, of natural language is a property that derives from *our* mental activity.

For Searle, this means that the intentional properties of computational properties are also derived from us. Searle contends that we have intentional properties not derived from anything, but we have them directly, since we have intrinsic intentionality. This intrinsic intentionality is what allows us to distinguish the meaning of "rabbit" from that of "undetached rabbit-parts," for instance (Searle 1987).

The realist conception Searle has about the semantics of the mental requires that the intentional features be intrinsic to the things that have them. He distinguishes metaphorical and derived intentional properties, which are not relevant to a realist conception of mind, from intrinsic intentional properties, which are essential to mind under a properly realist conception (Searle 1984).[17] To illustrate the distinction he uses the following examples:

A. Robert believes that Ronald Reagan is President.

B. "Es regent" means it's raining.

[17] Searle (2002, 63) speaks of observer independent, observer dependent, and metaphorical intentionality.

C. My car thermostat perceives changes in the engine temperature (Searle 1984, 3).

For Searle, if A is true, Robert's having the belief with the content, that Ronald Reagan is president, is an intrinsic intentional property of Robert. If B is true, "Es regent" has an extrinsic, or derived, intentional property with the content, that it is raining. Finally if C is true, the thermostat has intentional properties only metaphorically. He does concede, however, that while thermostats don't have real intentional properties, sentences of natural language do, even though such properties are extrinsically derived. Such intentional properties are, however, *assigned arbitrarily* by us to the intrinsically meaningless syntax of items like "es regent," "for that sentence might have meant something else or nothing at all" (Searle 1984, 5). Searle claims that like the meaning of "es regent," intentional properties of computational properties are not to be so much empirically found as personally imposed.

In Searle's conception this implies that a natural language sentence does not literally mean anything, though when we say that the literal meaning of "es regent" is that it is raining, we use a "shorthand for some statement or statements to the effect that speakers of German use the sentence literally to mean one thing rather than another" (Searle 1984, 4). Likewise, the symbolic character of properties a computational system is symbolic only insofar as we treat them that way. Searle remarks,

> If you open up your home computer, you are most unlikely to find any 0's and 1's or even a tape. But this does not really matter…To find out if an object is really a digital computer, it turns out that we do not actually have to look for 0's and 1's, etc.; we just have to look for something that we could treat as or count as or that could be used to function as 0's and 1's (1992, 839).

In contrast, Searle considers mental intentional properties to be intrinsic properties of reality, particularly biology, along with breathing, digesting, and sleeping.

"Intentional phenomena, like other biological phenomena, are real intrinsic features of certain biological organisms, in the same way that mitosis, meiosis, and the secretion of bile are real intrinsic features of biological organisms" (Searle 1984, 5). Searle considers all this intrinsic mental phenomena to be higher level phenomena with no more reason to be considered epiphenomena than other "higher level features of the world, such as the solidity of this typewriter" (Searle 1984, 6). Mental properties of mental phenomena are a part of the causal structure of the world, functioning "causally in the interactions between the organism and the rest of nature, and in the production of behavior. It is just a fact of biology that sometimes thirst will cause an organism to drink water, that hunger will cause it to seek and consume food, and that sexual desire will cause it to copulate" (Searle 1984, 10). Admittedly, humans engage in more sophisticated mental endeavors than other animals, where "the intentional state itself functions causally in the production of its own conditions of satisfaction or its conditions of satisfaction function causally in its production" (Searle 1984, 11), but this (supposed) difference does not cast it out of the biological realm. The problem is that, according to Searle, computational properties do not fit into this physical model (defined by intrinsic features), since they are "purely abstract" and are not in the physics as they "have no essential physical properties and *a fortiori* have no physical, causal properties" (Searle 1990, 116). Thus, Searle believes that computational properties are not intrinsic, empirically discoverable, nor causal, and therefore are not objective.

Searle: Computational Properties Are Purely Formal, and Therefore Not Objective

The second charge Searle makes is that computational properties are not objective because they are purely formal, and therefore not intrinsic, empirically discoverable, nor causal. According to Searle computation is purely formal, or “syntactic” (Searle 1990, 16). He expresses this as an axiom: Computer programs are purely formal (syntactic). Purely formal properties are not intrinsic, as they can be assigned arbitrarily, given the right number of parts.

> For any program and for any sufficiently complex object, there is some description of the object under which it is implementing the program. Thus for example the wall behind my back is right now implementing the Wordstar program, because there is some pattern of molecule movements that is isomorphic with the formal structure of Wordstar. But if the wall is implementing Wordstar, then if it is a big enough wall it is implementing any program, including any program implemented in the brain (Searle 1993, 841).

The problem with computation is that it is defined formally, or syntactically, and could therefore according to Searle, be defined over, rather than found in, any process with sufficient number of parts. Thus, any candidate computation putatively sufficient for cognition would be implemented on just about anything, including a wall or pail of water.

It follows, according to Searle, that computational properties, because they are formal, are not empirically discoverable nor causal. In contrast to the scientific understanding of light in terms of electromagnetism, which is a “causal story right down to the ground,” “the only power that symbols have, qua symbols, is the power to cause the next step in the program when the machine is running. And there is no question of waiting on further research to reveal the physical, causal properties of 0’s and 1's. The only relevant properties of 0's and 1's are abstract computational properties, and they are already well-known" (Searle 1990, 120). Computational properties are thus not objective in that they are not intrinsic, empirically discoverable, nor causal.

Searle: Computational Properties Are Multiply Realized and Therefore Not Objective

The third charge is that computational properties are not objective because they are multiply realizable, and therefore not intrinsic, empirically discoverable, nor causal. He claims that the "multiple realizability of computationally equivalent processes in different physical media is not just a sign that the processes are abstract, but that they are not intrinsic to the system at all" (Searle 1993, 842). He also claims that this lack of intrinsicality of multiply realizable computational processes rules them out from empirical discovery. "The aim of natural science is to discover and characterize features that are intrinsic to the natural world. By its own definitions of computation and cognition, there is no way that computational cognitive science could ever be a natural science, because computation is not an intrinsic feature of the world" (Searle 1993, 842). In fact, Searle sees computational cognitive science as deeply unscientific, endorsing the view that "the mind is completely independent of the brain" (Searle 1990, 121). Further, multiply realized computational properties are not causal in the way their realization bases are. For realization bases of actual mental computations are physical and governed by the powers of neurophysiology. Computations on the other hand, are not governed by such powers, since many kinds of physical properties with different physical powers could realize them. Thus, computations are not part of the causal world of the physical, as in order for a medium to compute “the physical features…are totally irrelevant” (Searle 1990, 121). In these ways, computation is not objective.

Searle: Computational Properties are Non-Objective Simulations

Searle’s fourth objection is that computational properties are not objective because they are at best simulations of objective properties, and therefore are not intrinsic, empirically discoverable, nor causal. The view that computational duplicates of mind are not intentional (or other constitutively mental phenomena) duplicates, but mere simulations is Weak AI. Searle endorses this view. Simulations of objective properties are simulations only insofar as the simulations have a relation to the thing simulated. The thought is not directly argued by Searle, but it is, without too much theoretical baggage, quite natural that simulations depend for their existence *as simulations* on what they are simulating. That is, you cannot have a simulation without anything simulated. The relation to the thing simulated admits of different theories, including relations to *possibilia* and abstract models, in order to account for simulating non-actual affairs. In both cases, however, simulations have essential extrinsic properties and therefore are not fully characterizable by reference merely to their intrinsic properties. Simulation properties are not intrinsic to the things that have them. Simulations are not empirically discoverable, as it is we who determine what it is that simulations simulate. What the thing simulated is is not to be found in the simulation itself. Searle contends that being such a simulation is not sufficient to have the causal powers of the thing simulated. As he famously put it, in a simulation of a thunderstorm no one gets wet. Thus, given that computational properties are at best as objective as simulation properties, they are not objective.

Non-Objective Insufficiency

Let’s remember the thesis of Non-Objective Insufficiency: Implementing non-objective properties is not sufficient for the implementation of objective properties. First, you don’t get from derived extrinsic intentionality alone, intrinsic intentionality. It is clear that Searle believes this from the remark that computers have at best derived extrinsic intentionality, and his belief that such intentionality could not support the putative intrinsic intentionality of computers. You don’t get non-formal elements from purely formal elements, or as he has it, syntax, being a non-objective property, is not sufficient for semantics, an objective property (Searle 1990, 117).[18] You don’t get non-multiply realizable properties from multiply realizable properties. He says, “the whole modern notion of computation that a program is multiply realizable in different hardwares, is sufficient to refute the claim that the implemented software by itself, regardless of the nature of the implementing medium, would be sufficient to guarantee the presence of mental contents” (Searle 2002, 52). And finally, you don’t get non-simulations, the objective standing of mental phenomena, from simulations, how computational cognitive science regards mental phenomena.

It follows that, given that the mental is intrinsic, non-formal, non-multiply realizable, and not a simulation (and is therefore allowed to be objective), implementing computational properties, which are, according to Searle, non-objective, is logically insufficient to implement mental properties. Or along another dimension, because the mental is intrinsic, empirically discoverable, and causal, whereas computational properties are not, implementing non-objective properties, which are not

[18] There is an interesting discussion of this idea in Rappaport (1994): if the syntactical properties alluded to by Searle are parsings, Rappaport contends that this is sufficient for self-understanding. Being embedded in the world, including with other speakers, is required only for mutual understanding.

intrinsic, do not admit of empirical discovery, nor do they cause anything, are they ever sufficient for the realization of mental properties by themselves.

1.3.4 Reasons Why Intentional Properties of Computers are Good Enough

Mental Realism is Common Ground

I will not criticize nor analyze the idea that mental properties are objective properties of machines much, as in this context it is fairly uncontroversial, and it is a premise both Searle and I accept. As becomes clear, I hold that my countenance of objective properties is just partly the countenance of facts in a particular field. If I sound like I've adopted Searle's framework it is only because I wish to engage with his view with as much understanding as possible, while ultimately undermining his conception. It is clear however, that computers are particular varieties of machines, and that there are machines that have different properties than computational machines, even within a mechanist worldview (Copeland 2000).

The Question of Intrinsic Intentionality in Computation

First, let us look at the charge that semantic (intentional) properties of computational properties are not objective because such semantic properties would be merely derived from us, and therefore are not intrinsic, empirically discoverable, or causal. What does it mean to say that a property is derived from us? According to Searle it means that the property is extrinsic and that its being a property of something else depends on the way mental beings treat it- what Searle terms observer-dependent.

Let us notice that a property's being extrinsic by itself should not make us believe that it is not objective. For example, as a Scientific Realist Searle should be comfortable with the idea that the property of the Earth, *being orbited by the moon*, is an objective property of it. So what notion of intrinsicness could be at work for him, since certainly he must allow that being orbited by the moon is an extrinsic property of the Earth, and is yet an objective property of it?

A conception of intrinsic properties is that it is those properties a thing could have if nothing else existed. However, Searle seems to be employing a close, but different notion. Being derived from us is being observer-dependent. The contrast class he uses for illustrating his notion of being derived from us is being intrinsic. "(T)here is a distinction between those things we might call *intrinsic* to nature and those features that exist *relative to the intentionality of observers, users, etc.* It is, for example, an intrinsic feature of the object in front of me that it has a certain mass and a certain chemical composition…But it is also true to say of that very object that it is a screwdriver. When I describe it as a screwdriver, I am specifying a feature of the object that is observer or user relative" (Searle 1995, 9). We may infer, as he does, that being intrinsic, in Searle's sense, is being observer-independent (Searle 2002, 62).

Searle thus claims that, in virtue solely of the computational properties of implemented computers, such computers will at best have derived intentionality, which is had by things only relative to our uses, thoughts, and observations, that is, observer-dependent intentionality. While our mental properties, particularly our consciousness and intentionality, are intrinsic, and thus not observer-dependent: "My present state of consciousness is entirely observer-independent. No matter what anybody thinks, I am now conscious. But the attributions that I make to my computer are observer-dependent" (Searle 2002, 62). For Searle to make his point, which is not about his current attributions of intentional properties to his current computer nor directly raising the question of consciousness, but rather the truth of the Computational Sufficiency Thesis, we must let the latter statement be rather about any possible attribution of intentionality to a possible implemented computer. Certainly, it is no “purely logical” argument against the Computational Sufficiency Thesis that because *Searle’s current attributions to his current computer* ascribe only derived intentionality, that no implemented computer would have the same objective kind of intentionality he has.

Given that Searle thinks that we have intentionality in the same way as we have consciousness, in that they are both intrinsic or observer-independent, he is committed to asserting the parallel statement that “No matter what anybody thinks, I am now thinking.”[19] Yet the claim has an immediate flavor of implausibility. Given that it is a prerequisite to his now thinking, *that he thinks*, how could his thinking be independent of anyone’s thinking, including his own? Could he be thinking without his thinking? There are several options that take Searle a bit further in trying to disentangle himself from this immediately apparent confusion, but not much I fear. Let us consider them. Hopefully the most plausible you see will be subsumed under one of them.

One option he has is to lay emphasis on the “what” in the “No matter *what* anybody thinks, I am now thinking.” Thus, he could plausibly claim that his thinking is independent of his thinking that he is thinking. Perhaps plausibly, he does not need to think that he is thinking in order to be thinking (although it might be controversial whether the thought that he is thinking is any implicit second-order thought attributable to Searle whenever he thinks). Further, perhaps he can even claim that his thinking is independent of his possibly actual thought *that he is not thinking*, under drugs, delusion,

[19] Searle thinks that intentionality and consciousness are intimately linked, but I will evade the complex issues raised by consciousness in order to carry out the discussion without undo complexity.

or a mental disorder (if such a thought is possible). But nevertheless, it does seem as if his thinking *something or other* determines that he thinks, and his thinking depends on his thinking the something or other that he thinks. For it is implausible that someone thinks, and yet thinks nothing at all. Unless, of course, Searle would like to countenance contentless or "merely formal" thought, which would of course defeat his purpose. So it is not true that his thinking is independent of what he thinks. In Searle's terms, his thinking, a supposed paradigm of an observer-independent property, is observer-dependent, and by his standards, not intrinsic.

The other option available to him is just to say that what he means by *anybody* in "No matter what anybody thinks, I am now thinking," has an implicit restriction in scope. The restriction is made explicit by the statement: "Independent of what *anybody else* thinks, I am thinking." In generalizing to a statement about the objectivity of thinking, one that disambiguates who the "I" is and can be endorsed by several speakers, there are two options. One is to say that "Independent of what anybody else thinks, Searle is thinking." If all that is what Searle wants to claim, then his view collapses into the view that an intrinsic or observer-independent property of the world is just a property that cannot depend on anybody but Searle. I don't think (objectively, I promise) Searle (and most people) will find this ontological egotism very plausible.

The other option Searle has is to say that "Independent of what anybody else thinks about who thinks, whoever thinks thinks." This option of course, allows that a machine that replicated our relevant computational properties would necessarily think, independent of what Searle thinks or anybody else thinks. Searle asserts, however, that it is impossible for a machine to have the intrinsic intentional properties we have in virtue of its computational properties, or in other words, that it is impossible for a computational machine, as such, to have the intentional properties it has without someone else thinking about it as having intentional properties.

Searle supports this by saying that whether a given computational property of a machine bears an intentional property is like deciding whether Calvin Coolidge was a great president of the United States, rather than whether Calvin Coolidge was born in the United States. Deciding whether Calvin Coolidge was born in the United States "can be ascertained as a matter of fact independent of the attitudes of the observers. But if I say 'Calvin Coolidge was a great President', that claim is epistemically subjective because its truth or falsity cannot be ascertained objectively, the claim can only be made relative to people's interests and evaluations" (Searle 2002, 66). To maintain the structure of the former proposition, Searle's contention is that Calvin Coolidge was born in the United States is observer-independent or intrinsic fact, expressed explicitly as "Calvin Coolidge was born in the United States independent of whatever anybody else thinks."

Here there are three points to be made. First, Calvin Coolidge's greatness can be determined by the facts. Surely, the extent to which the policies of his administration furthered US economic wellbeing, equality, freedom, democracy, and curved US imperialism determine the extent to which he was a great president. Such "property-clusters" associated with "greatness" are determined by the facts: namely the policies his government put in place and their effects. These are not matters of mere opinion.

Admittedly, there is an interest-relativity involved in *asking* whether Calvin Coolidge was a great president. But this brings us to the second point: that Calvin

Coolidge's being a great president or not involves an interest-relativity, or (to keep the terminology consistent) observer-dependency, that is also present in his being born in the United States or not. For the United States is not only a geographical place on Earth, it is an institution born around 1776 whose existence depends on people declaring and obtaining independence from England, crafting a constitution, and people treating a geographical location as the location of such an institution and treating each other as members of such an institution. The existence condition of the United States fits Searle's model of existence conditions for institutions. For example, that the geographical location where the United States in fact is, counts as United States territory, in the context of facts like those mentioned above has the logical structure of Searle's proposed institutional facts: "X counts as Y in context C."[20] So whether Calvin Coolidge was born in the United States depends on whether the territory in which he was born has the proper observer-relative context in which such a territory counts as national territory of the United States. Thus, it is simply not true that Calvin Coolidge was born in the United States is a fact independent of what anybody else thinks.

"OK, I might have been wrong about silent Cal's case" Searle might say, "but always having a negative charge is certainly an intrinsic or observer-independent property of electrons, while meaning that snow is white is a derived or observer-dependent property of "snow is white." Thus, I am able to maintain my distinction and say that our intentionality is intrinsic while a computer's is derived." This is the second way he indicates support for the idea that computational properties could not have intrinsic intentional properties.

[20] See Searle (1995, 80) particularly, but generally chapters 3-5.

Surely, there is something right when Searle says that it is a fact that all electrons have a negative charge independent of what people think. A likely source of the thought is that this fact, the fact to which we actually refer in the above expression, would remain in place were no minds to have ever said or thought anything about them, because the evolution of the universe just did not create the right conditions for minds to come about, for example. So (granting at least for the sake of argument) that while the *truths* about physics, chemistry, and (at least some) biology are observer-dependent (since we determine their content), the facts to which they refer are intrinsic. This is certainly right, but there is very little ontological profit to be gained from this assertion, if it is supposed to make a legitimate distinction between facts in physics, etc., and computational facts. All that means is that, *given the truth-conditions of* "electrons have a negative charge," the rest of the work required to make this statement true or false is done by electrons and their properties. The same applies to computational facts (there will be more on this later). Just think of a scenario where adders grew on trees and no other computational device did. Such adders, we might suppose, are just like your everyday calculator except that they just compute the PLUS function, without the rest of functions normally available. In that scenario "some adders grow on trees" would be true. Of course, if we meant *dividers* by "adders" the statement would be false. Likewise, *given the truth-conditions of* "adders grow on trees" the rest of the work required to make this statement true or false is done by adders and trees; and *the facts about adders would remain in place where we not here to talk about them*. In the same sense in which electrons always have a negative charge is an intrinsic fact, then, the fact that some adders grow on trees would be as well.

Perhaps some will regard as illegitimate to suppose that such fruits will compute the PLUS function, but not that electrons will have a negative charge, where there no minds around. They both are suppositions, but the very point is that there has not been said with sufficient persuasion WHY these facts are different, and why such fruits would not have their computational intentional properties (about addition, for example) unless we were born. Searle's repeating of his conclusion (or analogies incoherent with his own theory) is not independent justification.

In the absence of a successful justification for why such facts are so ontologically different, I regard them the same for simplicity's sake. We will consider possible justifications later in the chapter.

The Intrinsicness, Empirical Discoverability, and Causality of Natural Language Meaning

Searle urges us to use the model of derived intentionality exhibited by natural language utterances to model the intentionality computers have in virtue of the functions they compute. Thus, a question to ask is whether natural language intentionality is intrinsic, empirically discoverable, and causal.

Let us consider whether Searle intrinsically means rabbit by the natural language utterance “rabbit.” He says, “I know exactly which I mean,” rabbit or undetached rabbit part, by “rabbit,” “though someone might get it wrong about me, just as I might get it wrong about him” Searle 1987, 495). According Searle, this is a fact independent of what anybody else thinks, and therefore, by his standards, he intrinsically means rabbit by “rabbit.”

Can such meanings be empirically discovered? It is clear that when someone tries to learn a language there are semantic rules that govern utterances in that language that would make it inappropriate to say, “look “es regent” means that it is raining only derivedly, and therefore, it is not something I can discover about “es regent.” I can assign it any meaning I want.” I must confess I discovered something when I was informed that “es regent” means that it is raining. Further, anybody who has been to a country where they do not know the language knows that they fail to know something others in that country know. Linguists try to work out what correct translations for such languages are.

Suppose a body of Martian scientists came to Earth. The Martian chemist will find that Carbon is present on Earth. The Martian linguist will find that “es regent” means that it is raining. They can add these two bits of information to their body of discoveries about the character of things on Earth. They are on a par.

Further, by Searle’s standards, first-person judgments are sometimes just as good as other more paradigmatically empirical judgments. For example, Searle contends that through first-person judgments about our understanding of natural language utterances we come to know that there is no degenerately indeterminate semantics for natural language. He thinks there are facts about the meaning of words and what we mean by them, such that when we predicate about them, we get it right or we get it wrong. So there is something to be discovered about the semantics of natural language utterances in a way that is at least just as good as through other scientific empirical means.

It is also clear that the meaning of words not only plays a causal role that enables people to learn and understand the symbols of the language, but also in the use of the symbols of the language for all the socially coordinated (causal) tasks speakers of a language perform.

Thus, not only has Searle not given us good standards to say that the meaning of “es regent” or “rabbit” is not intrinsic, given the reason we have to say that it is (deriving from his own standards), he explicitly contradicts his idea that such natural language meaning is not intrinsic in arguing against Quine’s indeterminacy of translation thesis.[21] So if the semantics of computations are like the semantics of natural language utterances and Searle’s very idea of intrinsicness is of some consistent use, the semantics of computations are intrinsic.

1.3.5 Computational Properties are Not Purely Formal

Searle’s second reason for saying that computational properties are not objective properties of machines is that computational properties are purely formal, and therefore not intrinsic, empirically discoverable, nor causal.

Asking Whether Formal Properties are Intrinsic

Is it true that formal properties are not intrinsic? What notion of intrinsicness could be at play here? There is the notion of intrinsicness discussed in the previous section as observer-independent, and there is the notion of being not-extrinsic. But we know that the Earth has the non-intrinsic (extrinsic) formal property of being orbited by the moon in an elliptical shape, and that there should not be a problem with such a formal property being an extrinsic property or an observer-(in)dependent property of

[21] In fact Quine’s guess that there are no facts about meaning in his Indeterminacy of Translation thesis is not well supported. See Friedman (1975).

the Earth. For this formal property of the Earth is extrinsic, but is perfectly objective.[22] And, this formal property of the Earth is considered as observer-independent as any property figuring in the scientific realist picture of the world is. Further, there are also, of course, intrinsic formal properties of things. There is a standard sense in which shape is formal. Thus, for example, the property of being square is an intrinsic property, if any property is, and is formal as well. So it looks as though from the beginning Searle just does not have an argument.

There are delicate issues here, however. Searle's worry about the intrinsicness of formal properties can be expressed as follows. Computations are "purely formal (syntactic)" and therefore two formally equivalent systems are computationally equivalent. If the Computational Sufficiency Thesis is right, then any computationally equivalent systems are in an important sense psychologically equivalent.

The Sense of "Formal" Here is the Structuralist Sense

But Searle argues that being formally equivalent as a sufficient condition for being computationally equivalent and therefore psychologically equivalent (under the Computational Sufficiency Thesis) is either trivial or false. This is because, as he puts it, "syntax is not intrinsic to physics" (Searle 1993, 840). Rather, formal properties are properties of physical objects only insofar as we assign the parts of the abstract structure we conceive to parts of the structure of a target physical object, and this can be done to any object provided that it has the right number of parts. The way Searle expresses this, again, is:

[22] Perhaps to give him plausibility, we should give emphasis to Searle's link between intrinsicness (non-extrinsic) and nature, rather than intrinsicness (non-extrinsic) and particular objects as a condition on being observer-independent, as he sometimes suggests. In that case, *being orbited by the moon in an elliptical shape,* is an intrinsic feature in both senses of intrinsic.

> For any program and for any sufficiently complex object, there is some description of the object under which it is implementing the program. Thus for example the wall behind my back is right now implementing the Wordstar program, because there is some pattern of molecule movements that is isomorphic with the formal structure of Wordstar. But if the wall is implementing Wordstar, then if it is a big enough wall it is implementing any program, including any program implemented in the brain (Searle 1993, 841).

For Searle, any object with a sufficient number of parts has the same formal structure generally, and computer program particularly, as any conceivable structure or computer program. There are three comments to make at this juncture.

First, could cardinality itself be a constraint on what structure something has? How many objects exist in one set of four books? Two (couples)? 1,000 (pages)? 10^{29} (molecules)? Etc. Different sortals into which components of the aggregate falls into will give different cardinalities, and it will always be possible to construct the kind of sortal to break an aggregate into such that it has the desired cardinal number. You could divide an aggregate into portions of space it occupies. Given that each and every object can be arbitrarily divided into as many parts as there are numbers to divide the space they occupy by, and there are as a many of such numbers as any structure could require,[23] the "being sufficiently complex" requirement for structure identity does not rule out any object from having any putative structure.

In other words, the contended thesis here is that if it is true that any aggregate has any structure compatible with the same number of parts, then any object has any structure. That is to say that cardinality does not constrain what objects have the same structure. Call this the Useless Cardinal Thesis.

[23] This will surely be true for finite aggregates, which Russell thinks is "not improbable" to be the case in the world, in replying to Newman in a letter, quoted by Friedman and Demopoulos (1985, 631). The same is true if the structure is infinite, however, since we have available to us an infinite amount of point-particles to decompose any aggregate into and map the isomorphic structure onto it.

Yes, there is a question as to whether there are a given number of books or pages or molecules. The question is whether for any such *aggregate*, it has any particular structure rather than another and whether it cannot be said to have any structure. That is, the issue is whether it is itself redundant or trivial to say that the aggregate must have the right number of parts in order to have a given structure. Newman gives us the following definition of what it is for two aggregates, objects, or systems of relations (to give the notational variants in use) to have the same structure:

> Let a set, A, of objects be given, and a relation R which holds between certain subsets of A. Let B be a second set of objects, also provided with a relation S which holds between certain subsets of its members. The two systems are said to have the same structure if a (1, 1) correlation can be set up between the members of A and those of B such that if two members of A have relation R, their correlates have relation S, and vice versa (Newman 1928).

A set is a collection of objects; and we are asking whether sets with members have a particular number of objects within them. Let the set of two books mentioned before be such a system, “set A.” What cardinality does it have? It is correct that it has two *books*. But the question of interest here is, how many objects are in set A? Well pick any set of objects, say a set of twenty four dancing couples, “set B.” Does it have the same number of members as the set of two books? One can begin to address the question by thinking “no,” one has two members, while the other has twenty four. But it looks as though we can hold these collections of items fixed, specified in set-theoretic terminology if you like, and find without adding or subtracting anything from them they have the same number of objects. This is because we can say that the specified set of twenty four dancing couples just is identical with a set of two dozen dancing couples. In that case, the two sets turn out to have identical cardinality. This kind of individuation procedure is exactly what is required for declaring any two systems to have identical structure. Newman says that we assume a set of objects be given. The cardinality of those sets as such is not given. Rather, it is insofar as the members are typed in a certain way, that we get determinate answers to questions of cardinality. But there is no question of whether a set of objects, as objects, has a certain cardinality. Any cardinality can be said to hold of objects if they are typed in the right way, and the nature of the typing procedure for any desired case can always be found when one knows *ab initio* how many objects are desired. This is a way of always getting an aggregate to have a certain desired cardinality and therefore to have a certain structure. All of Newman's examples involve this procedure.

This is in line with his statement of structuralism that "All we can say is '*There is* a relation R such that the structure of the external world with reference to R is W'" (Newman 1928, 144). R determines how things are typed, and Newman considers that nothing of this relation is known except its existence. R can take any desired arbitrary value to yield the right cardinality characteristic of W.

There is nothing peculiar about the set of books and set of couples case. It is a mere example of the general phenomenon of being able to individuate any aggregate, or set, in such a way that it has the number of desired parts, or elements.

How wide is the scope of the consequent of the Useless Cardinal Thesis? I claim that it is exactly as wide as the scope of the antecedent- that is, they cover the same cases. The antecedent will cover finite aggregates, which Russell thinks is "not improbable" to be the case in the world.[24] Of course, this allows for the division of an aggregate into any number of parts as long as the number is finite. As we know, the number of finite numbers is infinite (e.g. there is no last finite number on the number line), though any definite finite number by which to divide the finite aggregate will yield a finite number of parts.

Perhaps the same is true if the structure is infinite, since we have available to us an infinite amount of point-particles to decompose any aggregate into and map the infinitely numbered isomorphic structure onto it.

There is a traditional question of whether matter is infinitely decomposable, and it is normal to interpret the question as being divided into two. That is, "Is matter decomposable into an infinite number of real particles?" and "Is matter composed of an infinite number of conceptually specifiable particles?" The first is supposed to be an empirical question, while the latter is to be settled *a priori.* But perhaps the breaking of the question into two in this way misses the point that what matters is the fact that as long as we can recognize infinitesimally small particles to be the bearers of a property, such as the having particular electromagnetic magnitude or the location of a centre, this is enough to vindicate their existence. This is the case with centres of gravity for example. They exist, they are average location of the weight of an object and they determine whether the object each of them is of is stable and how it moves. This is not akin to realism about non-existents. It is not awkward to claim non-existents don't exist, even if "realists" about non-existents will insist that there must be an existent for it to be the case that it has the property of non-existence. This line of thought I think just fails to understand the *negation* of existence. It is like arguing that since there are no

[24] Letter to Newman, quoted by Friedman and Demopoulos (1985, 631).

beers in the fridge there must be some kind of beer in the fridge to make this true. This makes for very thirsty drinkers. On the other hand, it is very awkward to say that objects do *not* have centres of gravity, given that it is a perfectly identifiable property of objects, upon which some of their movements depend.

Perhaps the same is true if the structure is infinite, since we have available to us an infinite amount of point-particles to decompose any aggregate into and map the infinitely numbered isomorphic structure onto it. Bringing in point-particles, however, creates the possibility of a singleton set of one point-particle. But since point-particles are partless, it looks like it would be a counterexample to the Useless Cardinal Thesis. This would be because only a singleton set of one point particle would be able to have such a structure. This case does not seem to be structurally identifiable with the set of two books, for example.

On the other hand, this case is strictly not covered by the definition of structural equivalence Newman gives precisely because it does not have proper parts and therefore is not susceptible to have the set containing it being divided into subsets, where a subset of a set, must not be the set itself but must leave a remainder.

On the other hand, if we agree, on the contrary, that a subset of the set can be the set itself, then, we are able to make the appropriate (1,1) correlations with the set of two sets of two books. For then we can give two independent properties to the point-particle. One set-membership condition that the point-particle we suppose satisfies is that it is my favorite point-particle. Call this set “Q.” Another set-membership condition that the point particle satisfies is that it is near my current location. Call this set “H.” Call the set that contains Q and H, “P.” Then, by the same standard that we allow Q and H to be subsets of P, we can call Q and H “parts of P.” Then, we've got two subsets, two members, two parts, given that parts do not have to be proper parts: we've got all we need for structural equivalence with the set of two books.

It should be noted that if these moves strike you as dodgy, it is because we allowed subsets to be the whole set. If we don't do this, then the point-particle case is outside the analysis of structure, requiring parts, that operates in the antecedent. Either way, the antecedent and the consequent cover the same cases.

The Useless Cardinal Thesis weakens Russell's structuralism even more. If his structuralist principles are upheld, not only can we not say that physics has intrinsic properties, is causal, nor is empirically discoverable: not even cardinality is part of the empirically discoverable.

Therefore cardinality is not a significant constraint either in the ways we may conspire to map abstract relational structures to aggregates. This position just strengthens rather than weakens Searle's argument though. Thus, in a sense it is too generous of Newman and perhaps Searle, to assert that cardinality is an objective property of aggregates, while structure is not.[25] The sense would be that in which it is given that Searle is successful in trivializing or falsifying computational cognitive science and Newman is successful in trivializing or falsifying a structuralist philosophy of what is known in the external world. However, Searle presupposes that

[25] Friedman and Demopoulos (1985, 627) are also too generous.

computational properties are purely formal in the way that structuralists claim the known physical world is. We shall see that this is an unsupported assumption.

A Double or No Standard at All

Secondly, Searle seems to be employing a double standard when he says that *because formal properties are not intrinsic to physics*, they are somehow too immanent (each is anywhere and everywhere!) to be of psychological or empirical importance or causal. But quite oddly, he does not seem to think the same applies to psychology even though he believes psychological facts are not to be found in physics- in other words, that psychology is not intrinsic to physics! For in arguing against Quine’s Indeterminacy of Translation Thesis, Searle agrees with Chomsky that “information at the level of microphysics is, by itself, not sufficient to determine the level of psychology… (I)f you fix the physics, the psychology is still open. For example, the theory of the dispositions of physical particles that go to make up my body, by itself, would leave open the question of whether I am in pain. The thesis that I am in pain is underdetermined at one remove” (Searle 1987, 487). This remark asserts that psychology is not fixed by physics, which is very puzzling because he imputes this thesis, as a reduction to absurdity, to people who endorse computational cognitive science, and says that this is anti-scientific and implies an untenable drawback to dualism (Searle 2002, 61).

But relevantly, if his contention is that psychology remains open once the physics is fixed, we may try to show an affinity between this view and computational cognitive science by saying that “if you fix the physics, the formal structure is still open, just as you would expect if you assume that if you fix the physics the psychology is still open.”

But perhaps the affinity disappears when you join it with the thesis that Searle holds, that “if you fix the physics, the formal structure is still open to whatever structure we would like assign it.” In this form we have a distinction between form and psychology, for according to Searle, form is observer-relative to whatever the observer wants to assign to an aggregate, and psychology is intrinsic. However, his contention does not hold much pressure, since it becomes a mystery why psychology is intrinsic to physics (given that being intrinsic to physics is the exhibited standard Searle favors), but is underdetermined by it, if being fixed by physics would be just what we would reasonably expect of anything intrinsic to physics.

Did It All Rest On An Equivocation?

Third, Searle attributes to computational cognitive science a structuralist philosophy of mind, one that could be found at least beginning with Russell’s more general structuralist philosophy of physics. Can we know anything about the physical world? And if so, what? Russell had a positive answer to the former question and a precise answer to the general latter question, presented most thoroughly in his *The Analysis of Matter*. His answer is instructive because of the affinity between Russell’s view of what we know of the physical world and what Searle says computational cognitive science tells us we know of mind, on the one hand, and between what Max Newman (1928) saw as a major fault in Russell’s theory and what Searle says is a major fault in the computational cognitive science theory. One way to put the matter is that Russell had a particular theory about the character of the knowable portions of the external world. The knowable portions of the external world are merely structural or formal. Given structuralism, computational cognitive science would show aspects of mind which are purely structural or formal. However, Newman showed Russell that his theory is fatally defective, as Russell admitted. Particularly, Russell's theory, Newman countered, implies that knowledge of the external world turns out to be non-intrinsic, not empirically discoverable but rather assigned or defined over, and not causal. This is just what Searle countenances against computational cognitive science. Russell responds in a letter,

> Dear Newman,
> Many thanks for sending me the off-print…I read it with great interest and some dismay. You make it entirely obvious that my statements to the effect that nothing is known about the physical world except its structure are either false or trivial, and I am somewhat ashamed at not having noticed the point for myself. It is of course obvious, as you point out, that the only effective assertion about the physical world involved in saying that it is susceptible to such and such a structure is an assertion about cardinal number.[26]

What was Russell's theory? Russell took it that we are not acquainted with the external world directly, for what we are acquainted with directly are merely our percepts. So what do we know of the things we take our percepts to be about, and what is the character of the knowable unobserved portions of the external world? According to Russell, our percepts preserve the structure of the external stimuli, while they can be grouped together with other internally structured percepts of other potential perceivers to yield "the conception of a space in which percipients are situated, and we find that in this space all the percepts belonging to one group (i.e. of the same physical object from the standpoint of commonsense) can be ordered about a centre, which we take to be where the object *is*" (Russell 1927, 216-217).

> For we assume that differences in percepts imply differences in stimuli…This principle, together with spatio-temporal continuity, suffices to give a great deal of knowledge as to the structure of stimuli. Their intrinsic characters, it is true, must remain unknown; but we may assume that stimuli causing us to hear notes

[26] Russell (1968, 176), quoted by Copeland (1996, 340) and Friedman and Demopoulos (1985).

> of different pitches form a series in respect of some character which corresponds causally with pitch, and we may make similar assumptions in regard to colour or any other character of sensations which is capable of serial arrangement…Nothing in physical science ever depends on actual qualities (Russell 1927, 226-227).

This results in the view that Newman eloquently expresses as: "The world consists of objects forming an aggregate whose structure with regard to a certain relation R is known, say (structure) W; but of the relation R nothing is known…but its existence. All we can say is '*There is* a relation R such that the structure of the external world with reference to R is W'" (Newman1928, 144). Thus, Russell's theory denies that even externally exemplified squareness or circularity are intrinsic to squares and circles, since these are properties of squares and circles merely with reference to an unknowable relation R. They are thus excluded by his lights from being intrinsic.

Key to understanding his thesis is to know what structure (form) is. According to Russell, two systems of relations, A and B, have the same structure if a (1,1) correlation can be set up between their members such that if two members of one system have a relation R their correlates in the other system have the relation S, and vice versa.

> For example, A might be a random collection of people, and R the two-termed-relation of being acquainted. A *map* of A can be made by making a dot on a piece of paper to represent each person, and joining with a line those pairs of dots which represent acquainted persons. Such a map is itself a system B, having the same structure as A, the generating relation being in this case being "joined by a line." The important feature of the definition, brought out by the example, is that it is not at all necessary for the objects composing A and B, nor the relations R and S, to be qualitatively similar… (their) intrinsic qualities are quite irrelevant (Newman 1928, 139).

The problem Newman points out and Russell recognizes is that "Any collection of things can be organized so as to have the structure W, provided that there are the right number of them" (Newman 1928, 144). Thus, at best, what is empirically discoverable is what the number of external things is. But this, as we saw before, would not really be a discovery, since the components of the structure it corresponds to are themselves divisible to provide any number required to have the same structure as could be said properly or improperly (guarding the distinction!) to be known.

Were what is known of the external world to be merely structural, fundamental questions, "to be answered by consideration of the evidence," would become a "matter of definition" (Newman 1928, 143), without significance. An example is the issue of whether "matter consists of atoms" is a true statement. Well, we could map (preserving structure) "matter" to the *London Underground Railway*, "atoms" to *stations* and "consists" to *in part consists*, and on that basis say that the "matter consists of atoms" is true. For any system, the same kind of labeling procedure could be performed and we are left with the insignificant matter of definition, rather than discovery, as to whether matter consists of atoms.[27] Newman recognizes this as a *reduction ad* absurdum of Russell's theory.

Could such a definition yield knowledge of causation? That is, if what is known is purely structural, could what is known be causal? What is particularly important for Russell in our knowledge of the physical world is our knowledge of causal relations, or what he calls "causal continuity." Causally continuous elements are those events represented by being near one another in a map. However, Newman countenances, if Russell's principles are upheld, this must be a mere definition. Were we to know something about its nature we would know something not structural (Newman 1928, 144-145). For, any pair of causally unrelated events can be mapped near one another to

[27] This seems to be Newman's (1928, 143) concern more explicitly developed.

be “causally continuous.” Thus, if Russell’s structuralist principles hold, we cannot say with any legitimacy that such and such an event is causally related to any other. What is known is not, on pain of triviality or implausibility since anything is causally related to anything, causal phenomena.

These two latter complaints are just what Searle argues is the third deficiency in the computationalist attempt to give an objective standing to mind, that is, that computational properties are not genuinely discoverable and that they are non-causal. His reason is that computational events as such are purely formal or structural, and this has the result that we can treat any entity as having the computational structure desired. The computational properties of a physical entity are true independent of how the physical entity is- that is, what properties it objectively has.

When is an entity said to compute a function? An initial intuitive account says,

> Entity *e* is computing function *f* if and only if there exists a labeling scheme L and a formal specification SPEC (of an architecture and an algorithm specific to the architecture that takes arguments of *f* as inputs and delivers values of *f* as outputs) such that (e,L) is model of SPEC.[28]

The problem with this definition is that it falls prey to what formalizes as *Searle’s Theorem:*

> For any entity x (with a sufficiently large number of discriminable parts) and for any architecture-algorithm specification y there exists a labeling scheme L such that (x,L) is a model of y (Copeland 1996, 338).

An architecture-algorithm specification is an exhaustive list of instructions calling “explicitly for the performance of some sequence of the primitive (or ‘atomic’) operations made available in the architecture” (Copeland 1996, 337). Briefly, the proof

[28] A sentence of the form “There exists a function *f* such that …” may be true or false independently of whether it is known to be true or false. Copeland uses the phrase “there exist a labeling scheme L and a formal specification SPEC such that…” in the same way.

consists in taking a standard architecture-algorithm specification of an entity M and mapping it onto an arbitrary entity which intuitively does not really compute the function in question (with sufficient number of parts), such as Searle's wall. This result is derived by obtaining a table of labels from a particular run of the program in question which may be given binary notation, and then designating these labels to regions in the history of the wall. Such labels correspond to moment-states of required registers and are interpreted, each, as "a function whose value at any point in the computation is the contents of the register at that point" (Copeland 1996, 344). Such a "Searlifying" labeling procedure will give the desired one-one correlation between a run of the program in question and moment-states of an entity which we otherwise say is of the wrong kind to compute the function in question- in this case, the wall. The wall is described by SPEC by interpreting its axioms in a particular way. Particularly, the interpretation will be a function whose domain is the set of moment-states of the particular run of the program and whose range is the set of moment-states of the wall. Such a procedure can be performed for any entity and any function to be computed, which yields Searle's Theorem as result (Copeland 1996, Section 4).

As Copeland points out, this result is valid only if a much more general thesis about the preservation of truth-value and subject-matter in theory, whether it is in physics, mathematics, geography, etc, is presumed. Formally, we may say that the grounding assumption is that a theory that is true under an intended interpretation will preserve its truth-value under a non-standard unintended interpretation. To illustrate, one could assign the number 1 as the referent of "London," the number 16 as the referent of "Moscow," and assign sentences of the form "a is north of b" truth-conditions of the form “the referent of “b” < the referent of “a.”” In this modeling “Moscow is north of London” is true, but no longer about Moscow and London (Copeland 1996, 447).

What is going on with Searle’s case of interest is that he is assigning non-standard interpretations to computational theories. We cannot reasonably expect to preserve truth-value in such a practice, since non-standard interpretations are a change of subject. Thus, where “the pair (e,L) is a model for some SPEC *but only under a nonstandard interpretation* of SPEC I will say that (e,L) is a nonstandard model of SPEC. A model that is not non-standard will be called *honest.”* Given that Searle’s Theorem trivializes the predicate “is computing the function *f*” Copeland proposes that computation is better conceived as:

> Entity *e* is computing function *f* if and only if there exist a labeling scheme L and a formal specification SPEC (of an architecture and an algorithm specific to the architecture that takes arguments of *f* as inputs and delivers values of *f* as outputs) such that (e,L) is an *honest* model of SPEC (Copeland 1996, 448).

Copeland thinks the wall is a non-standard model of SPEC for three reasons. First, SPEC is intended to be interpreted as a theory of computer action, but the wall did not contribute to the computation any more than a passive scoreboard contributes to a player’s move in billiards. Secondly, nonstandard interpretation introduces unintended temporal specificity into the theory, such that the wall is said to (non-standardly) compute but only for state-moments that have already occurred. The intended interpretation has implications as to how the entity having the computational properties would behave at times that outstrip the times to which the non-standard interpretation applies. Non-standard interpretations of SPEC are necessarily constructed *ex post facto,* and thus are always about the past, and therefore imply nothing about times for which the intended interpretation does. Thirdly, the conditionals expressing the algorithms in non-standard SPEC have the “force” of material implication, whereas the intended interpretation of SPEC’s conditionals (expressed by the primitive “If…ACTION-IS…”) is “stronger.” This can be seen in two ways. First, the intended interpretation of “If…ACTION-IS…” is that is it not true of an entity whose only credentials for satisfying it is that the antecedent is false. The non-standard interpretation of SPEC requires that this conditional be interpreted as a material implication, and under such a non-standard interpretation an entity satisfies the conditional when its antecedent is false. This has the undesired consequence that there is no distinction to be made between an entity that *does not* satisfy the antecedent but *would* satisfy the consequent if the antecedent was realized, on the one hand, and an entity that also *does not* satisfy the antecedent but *would not* satisfy the consequent if the antecedent was realized, on the other. Under the material implication interpretation, the conditional is equally true of both kinds of entities (Copeland 1996, 348-350).

Similarly, an entity whose only credentials for satisfying the conditional is that the consequent is true of it would not satisfy the strong conditionals but would satisfy the material implication analogue. It is important for the entity to satisfy the strong conditional rather than merely the material implication because it makes sense of the idea that the states of computing machine are “completely determined” by the antecedent states in a computational system, rather than acts of random chance that happen to satisfy the consequent.

Searle’s argument proceeds by equivocation, and while this may be permitted by structuralism, it should not permitted by the theory of computation. In fact, if Russell’s mathematical logic implies that equivocation is permitted, as we have seen, then logical truths like “A→A” would be considered as such even when the “A” in the antecedent means “coffee is black” and the “A” in the consequent means “water is made of gods.” But this is clearly a *reductio*. The logico-mathematical expression, to echo Russell, of the form of the antecedent and the consequent would be identical, expressed namely by “A.” But such shifts in meaning would not preserve the logical truth of the conditional “A→A.”

Further, if Russell's principles are upheld, his structuralist theory is equivalent to any other, putatively alternative, theory (with sufficient complexity, for whatever that is worth). For we could find one-one correlations between his theory and any other. If this is true, then his theory is really empty, in this other way as well.

There is one criticism of Copeland’s analysis that I would like to make. Copeland summarizes two necessary conditions for a model to be honest. One is that the conditionals specifying the machine’s behavior must secure the truth of appropriate counterfactuals. And secondly, the labeling scheme must not be *ex post facto* (Copeland 1996, 450). It seems to me like the force of this second requirement should be of some importance, but only secondary importance. It should have the importance accorded to the confirmation of theories constructed after the evidence is in. In general we should opt for theories whose test implications are true and are constructed *ab initio;* but this is not a necessary condition on a theory being true. Thus, theories confirmed by true test implications ascertained *after* theory construction are accorded more confirmation than theories confirmed by test implications ascertained to be true *before* theory construction. This rule of confirmation is to keep *us* honest and not pulling the sort of *ad hoc* maneuvers exemplified in non-standard interpretations. Nevertheless, a theory in general, and computational (empirical) theory in particular, cannot be refuted by reference to the fact that it is constructed after the test-implications of the theory are in-that is, after the run of the machine. It is a real possibility that a theory, including a computational labeling scheme, constructed *ex post facto* is, nevertheless, honestly modeled by an entity to which the labeling scheme is applied. In other words, a computational entity could make a theory true even if the theory was constructed after the entity performed a run or ceased to exist.

What I have said is already partly implicit in Copeland's intended interpretation of his definition of computation. As remarked before, Copeland intends his definition of computation to be interpreted in a particular way. A sentence of the form "There exists a function *f* such that …" may be true or false independently of whether it is known to be true or false. Copeland uses the phrase "there exist a labeling scheme and a formal specification such that…" in the same way (Copeland 1996, 338). This implies that (*e*,L) may be an honest model of SPEC independent of whether anyone actually specified L, and *a fortiori* (e,L) may be an honest model of SPEC independent of whether L was specified before or after *e* runs. I say that it is only partly explicit, only because as it applies to physical entities, "there exists a function *f* such that …" is true only if such a function is physically realized (independent of the character of Platonic functions if they exist).

Thus, Searle is just wrong if he thinks that programs are purely formal and that any entity is described by any specification of a program. It is surely true that labels that name states of physical devices are not intrinsic to those devices, if this is what Searle means by “syntax is not intrinsic to physics.” But this is not anything that does not attach to domains of any theory labeling portions of reality, including physical and causal theories. In fact, “electrons have a negative charge” is true under its standard interpretation, but false if it means that, for example, crises are never exploited for political gain. As was remarked before, statements about electrons and computations are on a par. They both require the physical world to be a particular way, and not just any way will do. Searle has not given us an argument why we should treat electrons and computations different. So (granting at least for the sake of argument) that while the *truths* about physics, chemistry, and (at least some) biology are observer-dependent (since we determine their content), the facts to which they refer are intrinsic. Given the standard truth-conditions of “electrons have a negative charge,” the rest of the work required to make this statement true or false is done by electrons and their properties. The same applies to computational facts. “Some adders grow on trees” is true if and only if some adders grow on trees.

Could There Be Computation Without Interpreters?

Let's suppose that there is some possible world *w* in which, by hypothesis, adders grow on trees. Adders, characteristically, compute the PLUS function. Further, let us suppose that in that world, there are no thinkers to have the intended interpretations of the kind required for us to say truthfully that our pocket calculator adds. It may be argued that such a scenario represents a counterexample to the computational theory considered here, because in that scenario there are no interpreters for such computational "fruits" to be honest models of, and yet, they are, by hypothesis, computational: they compute the PLUS function.

The computational theory considered here has resources to answer this objection. The theory can be used to point out that it is *our* computational theory with our intended interpretation that applies to these entities, and the intended interpretation of "in *w* entities growing on trees compute the PLUS function" is modeled by these entities: they, by hypothesis, satisfy the truth-conditions of the intended interpretation our statement "in *w* adders grow on trees" or its equivalent "in *w* some things that grow on trees compute the PLUS function." They satisfy the strong conditional counterfactuals characteristic of the PLUS function. The statements are therefore true. By the disquotational schema, that "p" is true if and only if p, these entities compute the PLUS function, even though there are no thinkers in that world to have intended interpretations of "in w adders grow on trees" or "in w some fruits compute the PLUS function." That such fruits grow in *w* does not depict a counterexample to the theory of computation explored here.

It may be responded that in a scenario where no one, ourselves included, had intended interpretations of "x computes the PLUS function" we would have to say, by our theory of computation, *that such entities do not compute the PLUS function*, even though such entities, by hypothesis, do. But this is a very peculiar response. It presupposes that we would have to give a truth-value to a statement, namely *that such* *entities compute the PLUS function*, without any intended interpretations and therefore without truth-conditions, which is absurd.

The kind of acknowledgement of ourserlves as meaning-makers, and therefore necessary for truth-makers to make our meanings true, inherent in this theory is not vicious and is quite general. It is not vicious because while it merely recognizes that mental beings exist and that one of our activities is to make meaning, and allows what our meanings are about to set their truth-value. In talking, for example, about electrons, it is they that make our assertions true or false. In talking about computations, it is they that make our assertions true or false. This generality holds until some good ground is given to the idea that somehow the facts about electrons and computations, those that determine a truth-value of our assertions, have relevantly different claims to existing independent of those assertions.

Syntactical Realism

Interestingly, structural realism or syntactical realism, as Worrall (1989) calls it, is just the view that Russell advocated regarding our knowledge of the external world. Worrall's motivations have more to do with the actual history of science than with an analysis of what percepts and theories can yield towards knowledge of the external world. Worrall considers that it is a requirement on any kind of realism that the history of science reveals that successor theories are cumulative with respect to ancestral theories. But the only thing that is seen to be preserved from theory to theory is their equations, *not the subject matter of the terms of the equations*, according to Worrall. From this analysis, it might be argued that unintended interpretations, yielding different referents for theory and successor, are a necessary component of the best realism the history of science can buy, and therefore, intended interpretations are not to be given the authority to determine whether a computational theory correctly applies to an entity or not.

The problem with this line of reasoning is that, first, Worrall’s structural realism is subject to the same reduction to absurdity that Newman showed applied to Russell’s theory of our knowledge of the external world. Anything can be shown to satisfy the equations of theory, given enough interpretational freedom. And secondly, Searle could not take the afore-mentioned line, since he is a full blooded Scientific Realist. It remains for Searle to show how computational cognitive science is committed to structuralism about mind, and how that way does not commit the rest of science, the bit he is wholly realist about, to the same doctrine.

What is required for Newman's and Searle's criticisms to hold is that Russell's theory and the theory of computation are merely formal. For it is only if this is the case that non-standard unintended interpretations of theories are permitted, which leads to bad consequences, as we have seen. However, while Russell explicitly committed himself to this theory, computational cognitive science does not. Thus, Searle’s attempt to show that material honest models of computation do not have their computational properties intrinsically, are not empirically discoverable, and are non-causal is unsuccessful. He therefore does not give good “formal” grounds for saying computational properties are not objective.

1.3.6 Multiple Realizations of the Same Objective Property

The third way in which Searle says that computational properties are not objective is from the fact that they are multiply realizable, and therefore not intrinsic, empirically discoverable, nor causal. Searle believes that the problems implicated in the computational theory of mind are already present in Turing's original expositions. Somewhat surprisingly, it is, at best, rare for Searle to show this with actual quotes and references, and the problem he sees with multiple realizability is a case in point. Here is a striking mention by Turing about what has come to be called "the multiple realizability" of computational states, with a connection to the multiple realizability of mental states:

> The storage was to be purely mechanical, using wheels and cards.
> The fact that Babbage's Analytical Engine was to be entirely mechanical will help us rid ourselves of a superstition. Importance is often attached to the fact that modern digital computers are electrical, and the nervous system is also electrical. Since Babbage's machine was not electrical, and since all digital computers are in a sense equivalent, we see that this use of electricity cannot be of theoretical importance… If we wish to find such similarities we should look rather for mathematical analogies of function (Turing 1950, 446).[29]

Searle believes that the apparent lack of interest, of the theory of computation in general and computational cognitive science in particular, in the material of the hardware implies that just any computational property is had by just any material hardware, thus robbing computational cognitive science of a purchase to objectivity. Searle believes that multiple realizability implies universal realizability, and universal realizability implies lack of intrinsicness, of empirical discoverability, and causal embeddedness, which in turn implies lack of objectivity.

But it is hard to see why universal realizability follows from multiple realizability. To take an analogy: Suppose one said that "there are an odd number of

[29] Alan Turing (1950, 446). Interestingly, Hilary Putnam is usually attributed the honour of introducing multiple realizability and computationalism to the philosophy of mind. See for example Bickle (2006): "In a series of papers published throughout the 1960s, Hilary Putnam introduced multiple realizability into the philosophy of mind." Surely Putman made significant contributions to understanding multiple realizability, but as so many other heretical theories Turing held, which then became central, the credit for introduction belongs to him. See also Turing (1948).

books in total on the table." There are many ways this statement would be made true. It would be made true by the table's having three, five, seven, and many more numbers of books. In fact there are, theoretically, an infinite number of numbers of books that could make the statement true. In this sense, a table's having an odd number of books is multiply realizable. In this case, theoretically at least, this property of the table is infinitely realizable. But this obviously does not imply that just any number of books on the table will be odd, making the statement above true.

Rigid Ships and Rigid Minds

Let us look at another case. The ship-kind is multiply realizable. Ships can be made of fiberglass, wood, and metal. In this sense, the property of being a ship is multiply realizable. A particular ship can be made of various particular materials. You can change a plank for another and still have the same ship, whether you change the plank or not. That is to say, whether a particular ship will continue to exist does not depend on your replacing one plank. In this sense, ships are token-multiply realizable, and the particular property of an object's being a ship is multiply realizable- that property is multiply realizable by the property of being made of certain particular planks or other. What is important about being a ship, and being the particular ship under consideration, is preserved under various multiple realizations of it.

One of the properties preserved when ships are the same ships under multiple realizations is rigidity. Rigidity itself- the power to not bend when a force is exerted on the thing that has it- is multiply realizable, since it can be realized in fiberglass and wood, but it is in an important sense the same property, exhibited to different degrees, in the multiple realizations of it.

The Searlean skeptic may try to draw from this story the apparent reduction to absurdity that the multiple realization thesis commits us to supposing that a ship’s fiberglass-realized rigidity is to be considered on a par with the rigidity of minds. There are minds that do not change their view by argument. Such minds are said to be rigid. The Searlean may impute the implication of the commitment to multiple realizability, that both the ship’s being made out of fiberglass, and a mind’s lack of responsiveness to reasons in a given subject, are both realizations of the same property: being rigid. It may be argued that given that having such diverse properties as being made out of fiberglass and lacking responsiveness to reasons can be realizations of the same property of rigidity, any similarly dissimilar properties can be realizations of identical properties. From this reasoning to universal realizability, Searle derives non-objectivity.

Important Properties

But it is clear that ships are not rigid in the same sense, for they are not similar in the same important way. Newman, and Demopoulos and Friedman, contend that this kind of way out posits the result that “’importance’ would have to reckoned among the prime unanalysable qualities of the constituents of the world, which is …, absurd.”[30] But there is nothing surprising nor troublesome about the idea that terms spelled the same but with different intended interpretations will consider for their satisfaction different properties to be relevant and therefore important, and others not. If it is required that “importance” is further analyzed so as not to be troublesome, then analyzing it in terms of “relevance for the satisfaction of the intended interpretation of the term in question,” will do, or so I think. Searle’s argument proceeds by the equivocation that structuralism allows, but is unjustified. Like rigidity or being odd,

[30] Friedman and Demopoulos (1985, 629) quoting Newman (1928, 147).

computational properties are not threatened from being non-intrinsic (in any non-general way, at least), empirically discoverable, nor non-causal. Thus, Computational Subjectivism is not supported by computational properties being multiply realizable.[31]

Searle's argument proceeds by equivocation, in assuming that the theory of computation can be given any non-standard interpretation, while retaining its standard content. While this may be permitted by structuralism, it is not permitted by the theory of computation standard understood. In fact, if Russell's mathematical logic implies that equivocation is permitted, as we have seen, then logical truths like "A→A" would be considered as such even when the "A" in the antecedent means "coffee is black" and the "A" in the consequent means "water is made of gods." This is clearly a *reduction* of Russell's structuralism, as was said before. It is unclear why Searle thinks the theory of computation is inherently structuralist, when it is clear that no theory should, and that to switch the standard truth-conditions of a theory for some other does not by itself undermine that the standard truth-conditions are satisfied (if they are).

Copeland's point is so basic that we can elucidate it like this. Picture a committed structuralist to revise *Searle's Theorem* like this:

> For any entity x (with a sufficiently large number of discriminable parts) and for any architecture-algorithm specification y there exists a labeling scheme L such that (x,L) is an honest model of y (Copeland 1996, 339).

[31] Searle (1993) contends that such a line of reasoning contrasts with neuroscientific evidence linking mental properties to quite specific neural structures. But of course, the claim that while some particular regions of the brain are particularly genetically determined to be pre-disposed to bear certain mental properties is compatible with the idea that such properties could be realized in silicon, for example. Further, evidence from neural plasticity already determines that specific regions of the brain do not necessarily bear the usual mental properties for which they are genetically predisposed. For example, Burton (2003), and Kujala *et al.* (2004) illustrate the different mental properties, including tactile and auditory, a particular region of the brain predisposed for visual processing can have.

We might think that this thesis is just clearly false, given that Copeland has introduced and defined the term “honest” such that it does not make just any entity (with the sufficient number of parts, for whatever that requirement is worth) an honest model of y. For instance, the wall will not satisfy the strong conditionals specified in SPEC. However, if our background theory is structuralism, we can map the interpretation of these strong counterfactual-supporting conditionals to weak material implication conditionals. We can do this by mapping the antecedent of the former to that of the latter and the consequent of the former to that of the latter; applying a function whose result is that the truth-value of any material implication about an entity is the truth-value of the strong conditional about that entity, and vice versa. Of course, “truth-value” and “strong conditionals” in use here do not have their standard intended interpretations; they have structuralism's correlates; call them “S-truth-value” and “S-strong conditionals.”

The theory of computation denies that this is legitimate only because, as Copeland makes explicit, putatively applying a theory equivocally is not really to apply it, but to apply some other theory using either the same words or some other subject-changing interpretations. Saying that models of computation must be honest is just a special case of saying that models of theories must satisfy the intended-interpretations of the theories that describe them. Non-standard models of theories are not real models of those theories any more than financial banks are models of the intended interpretation of “sides of rivers.” Further, the refutation of a non-standard interpretation of a theory is not really a refutation of the theory, which is what Searle does. Thus, Searle’s attempt to show that honest models of computation do not have their computational properties intrinsically, are not empirically discoverable, and are non-causal is unsuccessful.

Some may be uncomfortable with the idea of using the notion of “intended or standard interpretation” for saving computation from Searle’s charges, precisely because Searle would be put to rest in that computational facts are not independent of things like interpretations. According to the view expounded here, computation is independent in that computational facts would remain in place if no one did any interpretation. However, the genuine truths about computation, like the genuine truths of physics, depend on what we mean when we believe or assert them. For what we mean sets the conditions the facts must satisfy in order to make our beliefs and assertions true.

Thus, while we have not had to make changes to the Epistemic Thesis on this account, we have defended the approximate truth of a going paradigm in psychology against Searle’s threats. By the way we defended the Semantic Thesis we also see that that the theoretical entities of psychology are *bona fide,* since it was by refuting arbitrarily selective ontological skepticism about theoretical entities that we refuted selective semantic skepticism. Thus, in application to the mind, we can say the following:

4. ***The Epistemic Thesis about Mind:*** Current psychological theory is well-confirmed and approximately true of the world. So, entities posited by them, or at any rate, entities very similar to those posited, do inhabit the world.

1.4 Scientific Realism about the Mental

Scientific Realism and Mental Realism have effects on one another. Scientific Realism gives us confidence in the properties posited in our scientific investigations of the mind. Mental Realism refines our conception of Scientific Realism and constrains the avenues of its justification. With appropriate adaptations they happily co-exist.

2. Physicalism and the Problem of Mental Causation

The Problem of Mental Causation that originated with the contemporaries of Descartes arguing that his conception of the physical and his conception of the mental left no conceivable causal connection between things in those realms, has developed into the Problem of Causal Exclusion. This latter stage of the problem has been developed by Jaegwon Kim and threatens to undermine our *prima facie* justified belief in the existence of mental properties, events, and objects. The Problem of Causal Exclusion aims to show that if mental things (usually properties) are irreducibly mental, and Physicalism is true, there is no place for those properties to do irreducibly mental causal work. We can argue from this result that given that we should only be committed to the existence in nature of things that do causal work, or at least have a real potential do so (Alexander's dictum), we cannot be committed to the existence of mental things (including properties) in nature.

This Section of the dissertation begins by identifying the problem as formulated by Kim and looks at the positions putatively undermined by the argument, in Chapter 1. Chapter 2 proposes a solution to the problem based on an analysis of physicalism. This chapter leaves for a later one how the argument conceived as hinging on supervenience can be addressed. Chapter 3 critically analyses some key theses in Kim's development in writing about the Problem of Causal Exclusion. It is argued that while he has changed his mind on the issue of epiphenomenal causation, he has remained explicitly committed to three fundamentally inconsistent theses regarding higher-level, or non-basic, properties. Chapter 4 critically analyzes two important Non-Reductionist solutions to the Problem of Causal Exclusion, developed by Cynthia and Graham

Macdonald, on the one hand, and Derk Pereboom on the other. The Macdonalds argue for a version of the identity theory that construes mental and physical properties as distinct but instanced in identical events. Pereboom argues that the relation between the mental and the physical is that of constitution without identity. Chapter 5 tests some metaphysical aspects of the proposed solutions. We conclude that the problem is not a successful threat to Mental Realism in the last chapter, with a novel solution.

2.1 The Usurpation of Mentality as a Cause by the Physical

Different ways of formulating the problem of mental causation presuppose different conceptions of the physical. According to Jaegwon Kim, if physicalism is right then all properties either supervene on or are identical with physical property bases, and all properties that supervene on physical bases depend asymmetrically on their bases. Further, each supervenient property-instance is necessarily co-occurrent with its base. In either case, supervenient or identical properties occur simultaneously with their bases. If non-reductionism is right, then mental properties supervene on, but are not identical with, physical properties. The argument is premised on certain principles and those principles are applied to a particular case. The principles doing the work are the following:

> ***Supervenience*:** Mental properties supervene on physical 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).
>
> ***Exclusion*:** No single event can have more than one sufficient cause occurring at any given time- unless it is a genuine case of causal overdetermination (Kim 2005, 42).
>
> ***Closure*:** If a physical event has a cause that occurs at t, it has a sufficient physical cause that occurs at t (Kim 2005, 43).[32]

[32] I have added "sufficient" to the Closure principle because otherwise the argument does not get as far as being able to say that were there to be a mental cause of a physical event, it would have to be overdetermining, given that it *also* has a sufficient physical cause, which is claimed is not a viable option. The idea is to rule out the too-evident possible solution, without this sufficiency claim, that the mental is just a contributing, but not sufficient, cause, alongside the physical cause.

Suppose we have an instance of mental to mental causation. This would be a situation represented this way:

1. An instance of a mental property M1, occurring at t1, causes an instance of the mental property M2 occurring at t2.

It follows from our analysis of supervenience that:

2. The M2-instance has a physical supervenience base P2-instance, at t2.

A critical question arises: What made the M2-instance come about on this occasion? We get these possible answers:

3. M2 is instantiated on this occasion: a) because an instance of M1 caused an instance of M2, or b) because an instance of P2, the M2- instance's physical supervenience base, is instantiated.

The problem is that we have two competing explanations for the occurrence of this instance of M2, and P2 seems like an attractive candidate. Here's why: The occurrence of M2 is necessarily co-occurrent with one of its supervenience bases. At least in the circumstances, the P2-instance is sufficient for the M2-instance. The case seems like that the one involving a flashlight light circle moving on the wall. At no particular time does the lighted shape on the wall because the patio-temporally contiguous shape on the wall an instant later. What is doing all the work is the flashlight and its mover. Similarly, for Berkeley's conception of seemingly causal dependence between objects of perception, what is really doing the work is god's constant and regular creation, *ex nihilo,* of perceptual transmissions. Just like the spotlight on the wall now and the chair I see just now (on Berkeley's conception) did not depend at all on their prior confounding contiguous seeming "causes," the M2-

instance did not depend on the M1-instance. Let's try to vindicate the instance of M1's causing the instance of M2:

4. The M1-instance causes the M2-instance by causing the P2-instance.

Thus, 4. asserts that a supervenient mental property instance causes a subvenient physical property instance. This shows that: "Under the mind-body supervenience assumption, mental-to-mental causation implies, or presupposes, mental-to-physical causation" (Kim 1998, 47). So, the question now becomes whether we can make sense of mental to physical causation. From our analysis of supervenience we see that:

5. The M1-instance itself has a physical supervenience base, the P1-instance.

Now it looks like the P1-instance is going to preempt the M1-instance from causing the P2-instance. Since every physical instance that has a cause has a sufficient physical cause (by Closure), the P2-instance, being a physical instance that has a cause, has a physical cause- the P1-instance being the relevant candidate. What further causal work is there left for the M1-instance to do? It seems like the M1-instance's being a cause of the P2-instance, and therefore of the M2-instance, is dispensable.

The resulting alternative is:

6. The P1-instance causes the P2-instance and the M1-instance supervenes on the P1-instance and the M2-instance supervenes on the P2-instance. [33]

This yields the result that if the properties instantiated at a subvenient level are distinct from their resulting properties at a supervenient level, these latter properties are

[33] This argument can be found all over Kim's writings. However, I do not copy it word for word in each premise from any of these places, as he constantly flips back and forth in the premises, in using signs he at first seems to designate as standing for properties, like P's and M's, and then as property instances (including his 2005 work). This flipping back and forth is eliminated in the current presentation.

causally inoperative. What does all the work is exhaustively the physical subvenience base. It has the real causal powers of any instance of a supervenient property.

2.1.1 Functionalism is Vulnerable to the Causal Exclusion Argument

The metaphysical version of functionalism about the mind, what in the context of the philosophy of psychology (rather than physics) Ned Block (1980, 174) calls “Metaphysical Functionalism” is a commonly held theory of mental properties, and is assumed to be, at the very least, compatible with mental causation. Metaphysical Functionalism answers the question “What is it to be a mental?” by saying that the mental is functional. At the very minimum, functionalism assumes that certain kinds of explanations will apply to the things properly explained through the use of mental terms. Functionalism assumes that the behaviour of mental beings, such as persons, can be explained through the constituent parts of such beings and the way those parts are integrated together to generate the behaviour to be explained. Such an explanation would be akin to the explanation of how a factory generates ovens by relating the various assembly lines, workers, and machines composing the factory in question. Such a decomposition is what Robert Cummins (1975) calls “Functional Analysis.” Such an analysis explains the working of a given system and its capacities in terms of the working of the system’s parts and their capacities. It explains the functions of the whole system in terms of the functions of its parts. Some would like an account of intentionality, what Franz Brentano reified as the mark of the mental, in terms of a Functional Analysis (for example Cummins 1991 and Fodor 1992). Such an analysis is always relative to an analytical context about the nature of the system and its behaviour. For example “The heart functions as a pump” is correct given the analytical context of the circulatory system's ability to transport food, oxygen, and wastes. It is a contention against Cummins' account that different conceptions of the system and its parts will yield different functional analyses. Cummins's theory, for example, allows for saying that "the heart functions as a noise-maker" would be correct relative to a possible analytical context. It is contended that while this may be tolerable for certain things, it is not tolerable for an account of mental content, or intentionality. Intentional aspects of cognition seem to be about the things they are about quite determinately, and not just relative to our analytical presuppositions, which when varied, would shift what our mental states are about. In response some have adopted evolutionarily constrained interpretations, which purport to yield more determinate attributions of contents to mental states of thinkers (Prominently see Millikan 1990; 1993; 1996; 1999; 2008; Neander 2004; 2006; Shoemaker 1980; Tye 1997; Egan 2003).   How such "parts" and "their capacities" are to be understood is the fundamental matter upon which the two Metaphysical Functionalist theories disagree.

There are two theories commonly classed as Metaphysical Functionalist. One of these is Role-Functionalism and the other is Occupant-Functionalism. These theories are normally formulated in terms of properties, while being silent on whether mental objects and events are also functional. It is easy to see how a functionalist understanding of mental properties can be extended to objects and events, however, as functional objects and events would just be objects with functional properties and events that have or exemplify functional properties. According to Role-Functionalists there are *de re* identities between mental properties and functional properties, and between instances of mental properties and instances of functional properties. To say that such identities are *de re* is to say that there is nothing else to being a mental property or its instance but to be an irreducibly functional property or its instance. Functional properties and instances, according to Role-Functionalists, are by nature irreducibly relational in character. They are things that constitutively just have certain relations to other things. A role is constitutively something that mediates between two or more things. Hilary Putnam is commonly thought to be the original author of this view.[34] One way Putnam explicates how two systems have the same mental property is if the corresponding functional property of one system has a functional isomorphism to the other, such that "there is a correspondence between the states of one and the states of the other that preserves functional relations," Putnam (1967b) such as in computing machines.

Role Functionalism is sometimes defined as the view that mental properties identical with second-order functional properties. A second-order property is the property of having some property or other filling a causal role. The property filling the causal role is not however, the mental second-order property that specifies it. A stock example is the property of dormitivity. Dormitivity is the property of having some property or other that causes one to go to sleep when ingested. Such first-order properties can be the being diazepam or being liquor, for example, properties having the property of causing sleep and therefore being dormitive. According to the Role-Functionalists, any instance of being, say diazepam, also has a corresponding instance of being dormitive. Consider a pill of diazepam. It has the property of being dormitive; it is an instance of the property of being dormitive and of being diazepam; and it is an

[34] Block (1980) attributes this view to Putnam (1967a).

instance of being dormitive because it is an instance of being diazepam, which causes sleep.

In the scheme of Program Explanation developed by Frank Jackson and Philip Pettit, figuring in an explanation of why someone went to sleep, diazepam's being dormitive, is causally relevant but not causally efficacious (Jackson and Pettit 1990, 114).[35] Being describable as being dormitive, however, ensures that there is some property or other "programming" for the causation of the event to be explained.

> The property-instance does not figure in the productive process leading to the event but more or less ensures that a property-instance which is required for that process does figure. A useful metaphor for describing the role of the property is that its realization programs for the appearance of the productive property, and under a certain description, for the event produced. The analogy is with a computer program which ensures that certain things will happen- things satisfying certain descriptions- though all the work of producing those things goes on at a lower, mechanical level (Jackson and Pettit 1990, 115-116).

It is noteworthy that Jackson and Pettit are not only committed to the mental being epiphenomena, but see no problem in accepting properties that do no causal work. Program Explanation takes for granted that the Problem of Causal Exclusion has the right conclusion, even if such properties would be ruled out by Alexander's dictum. They do, however, share with the rest of Role Functionalists the assumption that mental terms are essential to the explanation of behaviour, and that to such terms correspond properties that are instantiated, or as they call it, "realized" in the world.

Occupant-Functionalists do not believe that having mental properties is irreducibly functional. Rather, having mental properties is reductively identical to having certain intrinsic physical properties, in the last analysis. This is what Ned Block

[35] At this point it is quite reasonable to contend, on the contrary, that properties can be causally relevant only if they are causally efficient as the Cynthia Macdonald and Graham Macdonald (2007).

calls “Functional Specification” Functionalism. David Lewis was a prominent advocate of this view.

David Lewis argues that the mental is functionally defined, and as we do with the rest of theoretical entities (which are all functionally defined), we reduce them to the intrinsic physical players of their functional roles (Lewis 1970; 1972; 1980; Armstrong 1970). Occupant-Functionalism just shares the ontology of the Mind-Body Identity Theory, the modern version of which was inherited from J.J.C. Smart’s 1962 “Sensations and Brain Processes.” The view is alluded to as a variety of Functionalism because it attributes certain relations to mental states, and then goes on to identify what has those relations. Such identifications are not posited independently, but *found* just like other theoretical identifications are found. The picture is like this:

Mental state M= the occupant of a causal role R (by definition of M).

Neural state N= the occupant causal role R (by the physiological theory).

Therefore, mental state M= neural state N (by transitivity of =) (Lewis 1972).

This account is exactly what we would expect if we were following a tradition of Scientific Realism that posited causal grounds of observed effects in properties intrinsic to the system in question. The mark of this account of Scientific Realism is that there are *discovered* identifications in science. However, by holding on to the Role-Functionalist conception, its advocates jeopardize this attractive apparent property of the Scientific Realist commitment (Pereboom 1991).

There is a slightly arbitrary issue with Lewis’s proposed property identification. It is that it is assumed that the occupant, at the type-level, will be essentially neurally specified. While this *may* be true, it is not determined by our existing state of knowledge, and it is not necessitated by the general conception Lewis advocates. For, we may define neural state N as well, and then find the occupant of the causal role in fundamental physics. The resulting derivation would look like this:

Neural state N= the occupant of a causal role R (by definition)[36]

Fundamental physics state P= the occupant of causal role R (by fundamental physical theory)

Neural state N= Fundamental physical state P (by transitivity of =)

This derivation would strike many as probably unsound, since physiological definition seems to be largely independent of the specific properties of fundamental physics. In this sense, the neural is irreducible with respect to fundamental physics. We may suppose that the same may go on for the mental, and just as the neural is not specifiable in terms of fundamental physics, neither is the mental. This does not mean that theoretical identifications for the case of the mental are not to be found. They may well be. But such identifications may well be mental.

The Role-Functionalist, but not the Occupant-Functionalist, is threatened by the Causal Exclusion Argument. For non-reductionists are interested in whether having irreducibly mental properties does causal work. According to the Occupant-Functionalist of Lewis's kind, mental properties do causal work only because they are reduced to neural properties. For Role-Functionalism the story is different. Since having functional role properties is supervenient on having certain intrinsic role-occupying properties (the role players), it follows that functional properties, by the

[36] Yes, a definition probably guided by other things we take ourselves to know, as Lewis allows, and as would be required for the definition of M as well.

argument as formulated above, are pre-empted from doing causal work by their realization bases, and therefore are not part of the causal structure of nature.

Let’s look at how the Problem of Causal Exclusion applies when we understand the mental as irreducibly functional. Let us apply it in terms of the role/occupant distinction, though it is clear that the argument could easily be formulated in terms of the first-order/ second-order distinction.

1’. An instance of a functional property M1 (M1-role), occurring at t1, causes an instance of the functional property M2 (M2-role) occurring at t2.

It follows from our analysis of functionalism that:

2’. The M2-role has a physical role-occupying base P2-instance (P2-occupant), at t2.

A critical question arises: What made the M2 come about on this occasion? We get these possible answers:

3’. M2-role is instantiated on this occasion: a) because the M1-role caused an M2-role, or b) because the P2-occupant, the M2’s physical occupant base, is instantiated.

Let’s try to vindicate the instance of M1’s causing the instance of M2:

4’. The M1-role causes the M2-role by causing the instantiation of the P2-occupant.

Thus, 4’. asserts that a role instance causes the instance of an occupant of a role. This shows that: “Under the mind-body supervenience assumption, mental-to-mental causation implies, or presupposes, mental-to-physical causation,” (Kim 1998, 43) or in current terms, mental to mental causation implies role-to-occupant causation. So, the question now becomes whether we can make sense of this kind of causation. From our analysis we see that:

5’. The M1-role itself has a physical P1-occupant.

Now it looks like the P1-instance is going to preempt the M1-instance from causing the P2-instance. Since every instanced role has an occupant of the role in question, what further causal work is there left for the M1-role to do? It seems as if the M1-role as a cause of the P2-occupant is pre-empted by the P1-occupant, and therefore the M1-instance is causally dispensable.

The resulting alternative is:

6’. The P1-occupant causes the P2-occupant and the M1-role supervenes on the P1-occupant and the M2-role supervenes on the P2-occupant.

This yields the result that if the properties instantiated at a subvenient level are distinct from their resulting properties at a supervenient level, these latter properties are causally inoperative. What does all the work is exhaustively the subvenient base. These constitute the real causal powers of any instance of apparent supervenient causes.

2.1.2 Anomalous Monism is Vulnerable to the Causal Exclusion Problem

Anomalous monism makes four fundamental claims. The first two have to do with the anomalous component of the view. They claim that there are no psychophysical laws connecting properties of the body and mind, and secondly that there are no mental causal laws. Thirdly, it asserts that while mental events exist, every mental event is a physical event, thus enforcing Physicalism, which is a position claiming that all falls within one category, a physical category. And fourthly, it asserts that there are causal interactions between individual mental and physical events.

The justification for the first claim is that "laws connect terms that we can tell *a priori* are suited to one another but psychophysical generalizations connect terms that we can tell *a priori* are unsuited to one another, so psychophysical generalizations cannot be laws" (Macdonald 1989; Davidson 1970). The way to tell *a priori* is by determining whether the terms which figure in psychophysical generalizations are homonomic or heteronomic. Generalizations express causal laws only if the terms employed are homonomic. Terms are homonomic if they employ concepts from the same conceptual repertoire. Otherwise, the terms are heteronomic, and cannot express causal laws. Macdonald and Davidson think that different constitutive principles from those that govern the ascription of mental predicates govern the ascription of physical predicates. These constitutive principles constrain the conditions of ascription of mental predicates to rational, consistent, and coherent beings, whereas they constrain physical predicates to things to which being rational, consistent, and coherent do not apply.[37] The result is that there can be no refinement in our conception of one domain by evidential considerations gathered in the other domain, for that would amount to changing the subject (Macdonald 1989, 88). Davidson says, "By changing the subject I mean here: deciding not to accept the criterion of the mental in terms of the vocabulary of propositional attitudes" (Davidson 1970, 112-113).

Several points are in order. Intentional psychology is a folk psychology, which we perhaps possess innately. But just as we understand "flat" space perhaps innately and can come to formalize it through Euclid's axioms, when we see that another framework does the job better we can change it. There is nothing holy about folk

[37] Nevertheless, Davidson needs to allow for degrees of irrationality and incoherence. Relevantly see Kornblith (1989, 207-214).

psychology that naturalistic inquiry is not able to challenge, revise or reinterpret. This is something Davidson would deny. Is it a change of subject if our fundamental conception of ourselves as rational beings changes? Well, in a sense, it might be said that Euclid and Riemann were talking about different things, but the relevant sense is the sense Newton hoped to capture: the fundamental structure of actual space. In this sense, Euclid's geometry was wrong and Riemann's seems to be right (even if we hold that Euclid's is right enough at the non-fundamental levels relevant for different areas of engineering for example). Scientific development is open to revision in theory, re-conception of the phenomena in question, and re-conception of the sentences used to describe the phenomena, and we should, on historical grounds, expect folk-psychology to undergo revision or re-conception. There are several ways this might go, including the possibility that our original conception is refuted.

We already see much re-conception of mind and mental properties as described by the possibly true theories of knowledge of language that Chomskyan linguistics offers. This scheme does not recognize the folk-psychological rational categories of belief and desire (nor does it refute it). Rather, the scheme recognizes other mental properties, like the having of an internalized grammar. It would be unjustified to argue that because internalized grammars do not have the categories of rational folk-psychology they are not mental.

Davidson's fundamental reason for thinking there are no mental causal laws is that causal laws must be strict, and putative mental causal laws would not be strict. Davidson's conception of strictness in law is of something "one could at best hope to find in a developed physics: a generalization that was not only law-like and true, but was as deterministic as nature can be found to be, was free from caveats and *ceteris paribus* clauses; that could, therefore, be viewed as treating the universe as a closed system." He argues that such laws could not cover "events when those events were described in the mental vocabulary" (Davidson 1993, 9).

Kim thinks that Anomalous Monism undermines irreducibly mental causation. According to Kim, the identity conditions of events are such that event (x, P, t) is identical with (y, Q, t') just in case x=y, P=Q, and t=t'. If mental event (Andrés, Sees that there is a bird, when Sun is ¾ down) = (Brain, Has visual cortex activation, 3 pm), then Andrés=Brain, Sees that there is a bird=Has visual cortex activation, and when Sun is ¾ down=3pm. Kim believes that were Andrés' seeing that there is a bird not reducible to Andrés' having a visual cortex activation, for whatever relevant physical supervenience base, Andrés' seeing that there is a bird would not make a causal difference. The consequence, then, is that mental causation is secured at the expense of irreducibility. Mental properties turn out to be physical properties.

Or conversely, irreducibility is secured at the expense of having the mental not make a causal difference. If we consider mental properties irreducibly distinct from physical properties, then because Physicalism is true and the strict laws of physics apply universally without exception, while mental properties are only anomalously exemplified, the Problem of Causal Exclusion emerges. If we suppose that any putative irreducibly mental cause has a physical effect, we will note that that physical effect is connected to a physical cause, upon which the putatively irreducible mental cause supervenes, exemplifying a strict causal law sufficient to bring about the effect in question. But given that the putatively mental cause is superveniently dependent on its physical base, and given that its connection to the effect is not bound by the strict laws that bind causes and effects, irreducibly mental causes are excluded from participating in the causal structure of the physical world by the physical properties of the subvenient physical cause.

Let us look at Problem of Causal Exclusion bit by bit under the understanding of Anomalous Monism.

1*. An instance of an anomalous property M1, occurring at t1, causes an instance of the anomalous property M2 occurring at t2.

Davidson thinks that all actual events and their causal connections are backed by strict laws linking properties:

2*. The M2-instance is causally backed by the (strict) lawfully-integrated base P2-instance, at t2.

A critical question arises: What made the M2 come about on this occasion? We get these possible answers:

3*. The M2-property is instantiated on this occasion: a) because the anomalous M1-instance caused an M2-instance, or b) because the (strictly) lawful P2-instance, M2's physical causal-integration base, is instantiated.

Let's try to vindicate the instance of M1's causing the instance of M2:

4*. The M1-instance causes the M2-instance by causing the instantiation of the P2-instance.

Thus, 4. asserts that an anomalous instance causes a lawful instance. This shows that: Under mind-body Anomalous Monism mental to mental causation presupposes that an instance of a property not lawfully integrated into the causal structure of nature causes an instance of a property that is lawfully integrated into the causal structure of nature. So, the question now becomes whether we can make sense of such anomalous to lawful causation.

5*. The M1-instance itself has a (strictly) lawfully integrated P1 realizer.

Now it looks like the P1-instance is going to preempt the M1-instance from causing the P2-instance. Since every lawfully related instantiation of a property that has a cause is necessitated by the law the property is related by, what further causal work is there left for an anomalous M1-instance to do? It seems like the M1-instance as a cause of the P2-instance is pre-empted by the P1-instance, and therefore the M2-instance is causally dispensable.

The resulting alternative is:

6*. The P1-instance causes the P2-instance and the anomalous M1-instance supervenes on the P1-instance and the M2-instance supervenes on the P2-instance.

This yields the result that if the properties instantiated at a strictly lawful subvenient level are distinct from their resulting properties at a supervenient anomalous level, these latter properties are causally inoperative. What does all the work is exhaustively the subvenient base. The base has the real causal powers of any instance of apparent supervenient causes. According to the conclusion Kim draws, Anomalous Monism implies epiphenomenalism.[38]

[38] It should be pointed out that this is not what Davidson thinks, because he does not accept the employment of properties in this way; he talks of predicates and events, not properties, so the epiphenomenalism does not follow for him.

2.2 The Argumentative Genesis and the "Physical Winner"

The Problem of Causal Exclusion descends from the causal problem Descartes' contemporaries posed to his Substance Dualism.[39] The arguments of Princess Elisabeth of Bohemia and Pierre Gassendi are well known. The traditional historical view is that Descartes could not adequately answer the question generated by his conception of the physical and the mental, of how it could possibly be that mind outside space could causally influence, and be influenced by, the body located in space. Physicalist monism was a natural alternative that seemed not to be vulnerable to the objections Princess Elisabeth and Gassendi posed.

One commentator says that philosophers essentially got things wrong when they "arrested" Descartes' Dualist doctrine on the basis of causal considerations. Louis Loeb (1981) argues that all that Descartes needs to answer the question posed by his critics is a Humean conception of causation. Armed with such a "constant conjunctionist" conception Descartes could just say that the causal connection was fundamental, and that at some level we all had to accept that there are no further explanations for causal

[39] In presenting his argument for what is normally taken to be Substance Dualism Descartes seems go from premises asserting that his mind and body have different properties to the conclusion that his mind and body are distinct. However, mind-body non-identity is a much weaker than what a Substance Dualism worth asserting properly requires. For while it might be correct to say that a statue and lump of clay are distinct, since they have different properties, this would not be a proper ground for Substance Dualism about statues and lumps of clay. In light of such considerations, Yablo (1990) reconstructs what a proper argument for Substance Dualism is supposed to be like. In order to rule out the distinctness of lump of clay and statue as a basis for Substance Dualism, Yablo says that Substance Dualism contends that A.) that all the mind's "intrinsic, categorical, properties are mental rather than physical" and B.) all of the body's "intrinsic, categorical, properties are physical rather than mental." He concludes from this that Substance Dualism asserts that in actuality minds are in all intrinsic and categorical respects indiscernible from pure disembodiedment (153). Now disembodiedment should not be understood merely as being unextended. For space-time points are unextended, but locatable and physical. It is therefore open that, on a par with space-time points, categorical and intrinsic properties of mind are mental and yet lawfully within space-time. In that case, we would still have no proper conception of Substance Dualism. How then to understand "being disembodied in all intrinsic and categorical respects"? A viable answer is that in all intrinsic and categorical respects minds are outside of space. Bodies on the other hand are in all intrinsic and categorical respects within spacetime.

connections. As Hume argued, there need not be any likeness between cause and effect from which one can infer there to be a necessary connection; the causal connection is a non-conceptual connection.

Kim believes Loeb is wrong and his reasons are instructive. Kim thinks that such a conception of causation between mental events and physical events, fails to recognize the difference between two agents psycho-physically synchronized in such a way that all their intrinsic mental and physical properties were indiscernible. Take the case of Smith and Jones who are psycho-physically synchronized to will their hands to rise, and their hands rise. Now according the Constant Conjunction theory the following possibilities are not ruled out. Smith's will caused Jones's hand to rise and Jone's will caused Smith's hand to rise; Smith and Jones' wills together (overdeterminedly) caused each of their hands to rise; and Smith's and Jones' wills caused their hands to rise together, each contributing to the other's. Clearly these options must be wrong. The same problem arises if the effect is mental, that is, if the putative case involves mental to mental causation. The problem illustrated by the examples is what Kim calls "the pairing problem," that is, the problem of how to couple genuine causes to their genuine effects. On Kim's diagnosis what creates the problem is that at least one of the relata of the causal relations is not spatially contiguous with the other. "The temporal order alone will not be sufficient as a causal framework for this purpose. It was not for nothing that Hume included "contiguity" in space and time, as well as constant conjunction and temporal precedence," (Kim 2005, 86) in his analysis of causation. Armed with the addition of spatial contiguity, we get the right answer to the question of whose will caused whose arm to rise. Kim believes this is the only way of securing some hold on causation that yields the correct answer. He holds that mental analogues of such a physical space framework are plainly inconceivable. “But I don’t think we have any idea what such a framework might look like- what purely psychological relations might generate such a space-like structure. I don’t think we have any idea where to begin,” (Kim 2005, 82). A particular challenge he poses is to find a place for the relation of *being between* that is not couched on the physical space framework. In sum, we can take Kim’s causal reasons against Substance Dualism to be twofold: 1. That the only way to solve the pairing problem involves locating the causes of action spatially contiguous with the action in question. And 2. That we cannot even begin to get an idea on how causation outside of (physical) space might even happen- there being no other, to his mind, framework generating such connections.

My position, to state it before arguing for it, is that I disagree with Kim that the only way to solve the pairing problem is by locating causes and effects in physically contiguous space, and that non-physical causal frameworks *cannot* be devised. I agree with him though that realist laws are ineliminable- my conception of laws asserts with the Humean that at the end of the road they’ve got to be brute, i.e. fundamental—but like Kim, I think they are not brute constant conjunctionist. This allows for the *formulation* of Kim’s pairing problem in the first place, and allows for the *possibility* of mental causation under Substance Dualism- but this does not mean that it becomes deeply underdetermined which of Kim’s doppelganger alternative putative causes are the real causes.

I believe Kim is not right in asserting that spatial contiguity is necessary for causation. Kim supposes that spatial existence is necessary for causal connections to exist, grounded on the idea, taken as fact, that spatial contiguity is necessary for causal connections to exist- given, as he takes it, that this is the only way to solve the pairing problem. Thus, if we challenge the assumption that spatial contiguity is necessary for causal connections, then we challenge the idea that spatial existence is necessary for causation. Newton designed a system under which causation at a distance was possible and actual, for instance. His theory of gravitation presupposed breaking Kim's stricture that causes and effects be spatially contiguously located. It was surely open, however, for Newton's challengers to criticize his system based on the pairing problem. They could have argued that worlds in which our solar system was perfectly duplicated, that is, in which our solar system was synchronized with a doppelganger, it was open that, to emulate one of Kim's options, their doppelganger Sun was attracting our planet to our Sun and our Sun was attracting their doppelganger planet to their doppelganger Sun- his equations would describe the system's observed behaviour equally well. Thought experiments have their place in our knowledge-seeking practices. There are, however some considerations to weigh against how much this particular thought experiment, derived from the pairing problem, should move us in believing the impossibility of psycho-physical and pure mental to mental causation under Substance Dualism.

It is generally accepted that Newton's paradigm was a definite success with respect to other contemporaneous contenders, and therefore was a good theory to accept at the time, regardless of the possibility generated by Newtonian action at a distance that the real causes of the observed effects we tried to explain were some synchronized duplicates which exerted their force at a distance. Surely, this troublesome possibility arises for the constant conjunctionist conception of causation in general, and the realist conception of causation can do no more than assert that one of the putative synchronized causes is the real cause in the considered cases. We would all want to assert that the relations of dependency described by Newton equation for gravitational attraction is a law that *regulates* the behaviour of *our* Sun and planets- not that it obtains as matter of *coincidence,* where the highly unlikely synchronized doppelgangers are the real sources of causal influence on our planets. But it is sufficient to ward off this doppelganger-interaction possibility as actually obtaining by using other tools of scientific rationality. By these methods, simplicity rules out synchronized duplicates as doing the causal work, in recognition of the wanted verdict, that the gravitational laws connect the Sun and the planets *in virtue of* their masses and locations with respect to one another. Considerations of parsimonious predictive success are a tenet of scientific rationality. Positing such doppelganger-infested scenarios as hypotheses obviously complicates the picture beyond justification, without generating some surplus in predictive or theoretical success. We can call this the Reply from Plain and Simple Ockham. It by no means blocks the possibility, but it is cast as a "far away" possibility not worthy our belief, as it should be.

As a historical point, we might add, what moved the scientific community to accept a new paradigm was more to do with the constraints imposed by scientific rationality rather than the theoretical possibility that Newton's gravitational equations described coincidental relations between the planets and Sun, whose real causal ground lay in the putative law existing between our planets and doppelganger Sun. Again, this is done without the need for contiguity in space, so it is not the case that spatial contiguity is an ineliminable part of the solution to the pairing problem.

It may be thought that the conclusion outruns the argument in an illegitimate way. It might be argued that “as it stands it looks like an inductive argument saying that in this case, we can differentiate between real and only apparent causation without the need for spatial contiguity, and therefore we do not (ever) need spatial contiguity to differentiate between them (real and only apparent causation). But you have relied on simplicity, so it is still an open possibility that there may well be ways of constructing a system of real and apparent causal connections which are equally simple– the equations describing the two putative causal chains are much the same in terms of simplicity, and that the only way of distinguishing real from apparent is via contiguity.” But Kim’s argument cannot be that there just may be, in the abstract, a case where we have two non-identical competing scientific theories with equal evidential and theoretical support, including considerations of simplicity. This would just be an uninteresting underdetermination thesis without much motivation and I assume, as Kim does as well, Scientific Realism, which implies that there are good criteria for theory selection. Further, the doppelganger case of our solar system share’s the fundamental features of the “doppelganger wills” case Kim uses to motivate the necessity of spatial contiguity: that they are synchronized doppelganger duplicates. The analogy is a good one, I think, and the induction is *bona fide.* Regardless of whether, in principle, it is possible to generate empirical theories that are equal from an evidential and theoretical point of view, the case Kim uses to show the truth of his (inductive) conclusion, that whenever we posit non-local causation the case will be underdetermined and therefore that all causation must be local, does not do the work, as it can be ruled out by other means. Since, then, no compelling reason has been given for the necessity of local causation, we should not accept it.

Today we find the phenomena of entanglement, where properties of some observed objects depend on properties of other observed objects at some other location.[40] It is simply an open question, from an *a priori* perspective, whether action at a distance is occurring. The question is solved through *a posteriori* methods, and by these methods, it looks like action at a distance is *required* rather than prohibited. If we need action at a distance to explain what we need to explain, if that is our best theory of the world, then we are justified in believing in it.

Further, it is a *possibility* that there are laws determining the movement of our arms, as well as the movements of the planets, which connect them not to what we believe are their causes, but to some synchronized doppelganger. I am, after all, a fallibilist about my beliefs and liberal about what is logically possible. For this reason, I also believe it is a real possibility that purely mental events in the "second substance" of Substance Dualism cause other such events and events in the physical world. This, however, does not mean that we cannot make a justified judgment about which possibility is actual, based on the already existing tenets of scientific rationality (parsimony being a relevant example).

How are we to pair each of the synchronized causes to their synchronized effects? It is certainly *possible* that Jones' will causes Smith's arm to rise, and vice versa, or any other "pairing problem" scenarios. His example is supposed to give us the

[40] If entanglement is a *sui generis* non-causal relation because it involves the transfer of information faster than the speed of light, to which all causation is constrained, then take your pick: particles that go in and out of existence between locations that are not contiguous are an obvious alternative.

intuition that we *must* be wrong if we think this actually happened. But I think this is perfectly possible and thus realized in some distant possible world where there is a Smith and Jones who, unbeknownst to them, synchronically control each other's arms but not their own. This should not be surprising, given the bizarre worlds that possibly exist, and the extremely bizarre world Kim has picked- where such improbable synchronies exist. For the rest of us, we might go and test whether people's wills attach to their movings or to others' in the normal naturalistically fallible, but fruitful and often knowledge-conducive way. In this way we may find that people's wills do in fact cause their hands to rise. If Kim then objects that our method does not rule out that the people we have tested all have synchronized twins, whose wills are the real causes of their risings Kim himself has just presupposed action at distance of a kind he abhors by the wills of almost impossible characters (provided doppelganger twins don't have wills flowing out of their bodies and into ours, billions of kilometres away, faster than the speed of light).

So this consideration shows two things: that it is sufficient to wield the tenets of scientific rationality in order to ground verdicts about the alternative causal sources of phenomena brought to discussion by the pairing problem. And secondly, that scientific rationality does not rule out laws that govern events happening at a distance. This second conclusion is strengthened by today's physical science.

One base for Kim's thought that all elements of causation must be in space is that causes need to be spatially contiguous with their effects (in order to solve the pairing problem). Since this base does not stand up to pressure, there is only the idea that we cannot even begin to conceive of non-spatial causation- causation whose elements are not in space- to support his thesis.

For the second route: the “no idea where to begin” argument, I think Kim is being very uninventive when he says that we don’t have any idea of where to begin to build a model of a purely mental realm that provides us with items that are *between*. Clearly colour-space provides a place for colours between others with respect to their brightness and hue for example. True, this is a “space of a sort.” But the relevant space Kim alludes to is physical space, the one that can be measured with rulers, contains depth, height, and length, etc. The dimensions of colour-space are not of the same kind, and therefore the “space” represented by them is not the same thing Kim refers to as necessary for causation. The dimensions in a Cartesian graph for colour-space can represent hue, saturation, and brightness. It might be said that we need a way of making causation evident in this Cartesian graph representation. In that case we can add a temporal dimension to the graph, while fixing colour properties represented on both of the other axes, yielding different combinatorial results. The result is a representation of how colours interact when mixed in different proportions and representations of possible instances can be plotted at will. While Cartesian graphs have been employed to represent physical space, they need not be. Their axes can be used to represent many other possible magnitudes, and their contents can represent “places” in colour-space (rather than physical space). The functions that would legislate the way some properties of colour fix others would be descriptions of laws that obtain in such pure colour worlds. Different elements on the graph will find elements between them, so this approach does give us a place for *being in between*.

What really should motivate Physicalism is not, I think, Descartes' inability to say just how mental and physical events of minds and bodies could causally interact. Rather, what should motivate it is a particular vision of how science develops and unifies domains of inquiry. For example, how chemistry unifies parts of biology in biochemistry, and physics unifies parts of chemistry in quantum chemistry. It should come as no surprise that mental properties are properties of organized matter, any more than repulsion and attraction are properties of matter, and that God might have "super-added to matter a faculty of thinking" just as he "annexed effects to motion that we can in no way conceive motion able to produce," (Locke 1690, 540) or more convincingly, just as he annexed the power to exercise actions from a distance, or other modern examples.

2.2.1 Unpacking Physicalism

A very peculiar state of affairs arises after the demise of Contact Mechanics, and the allowance of "occult qualities" to count as physical. While "the physical" was a well defined notion for Descartes, it may not be a well a defined notion now. For it is unclear under what conditions Physicalism is true or false. Chomsky has been critical of the idea that we have the right to take for granted that we know what the physical world is like and that we have some definite conception of the physical (or mental for that matter), that enables us to carry out metaphysical debates about the materiality or not of mind. Chomsky thinks that methodologically, the assumption that naturalistic inquiry into language needs to explain it in physicalistic terms has not played a role in the development of psycho-linguistics. We do not appeal to atom nuclei, for example, in search of understanding knowledge of language. In fact, it was the assumption that we can tackle problems of language and mind independent of problems of physics that has made possible the current scientific understanding of language and mind (Chomsky 2000, Chap. 4).

However, neurons are theoretical entities that do not come from the theoretical universe of physics. They are biological entities, as minds are psychological entities, which also don't come from the theoretical universe of physics. The argumentative genesis of Descartes' Substance Dualism appears to have fed the reproduction among the philosophical population of the idea that all is physical, in chemistry, and biology, but not the mind. It is unclear why minds cannot be as physical as neurons are accepted to be.

It is reasonable for physicalists to want a conception of the physical with truth-valuable content, that is not trivial and yet has enough flexibility to allow for the multiplicity of posits in historical scientific and philosophical development, that is not refuted by current physics, that would not count immaterial things as physical, and that is not self-defeating. There are several proposals competing in the philosophical market today, but I think that they succumb for failure to satisfy the theoretical desiderata specified above. Of course, if no other proposal could satisfy all these desiderata we might settle for the least bad, if not optimal, option. However, I do think a proposal that satisfies all these desiderata exists. In the following pages I will provide a critique of existing conceptions of the physical that may be used to give content to physicalism- fundamentally, the idea that all is physical- based on their failure to satisfy the mentioned desiderata. I will then present my own conception, one that I will argue does not fall prey to the criticisms that undermine the others and has several additional virtues.

2.2.2 Chomsky's Challenge to Physicalism and the Inadequacy of Physical Spiritualism

The typical physicalist thesis looks like this:

Physicalism: A minimal physical duplicate of our world is a duplicate *simpliciter*.[41]

This thesis implies that once you set the distribution of physical entities as they are distributed in our actual world, without remainder, you set the distribution of all other phenomena, including mental phenomena, for example. There is no physical duplicate of our world, without remainder, that does not duplicate the mental phenomena of our world, if Physicalism is true.

However, the content of Physicalism depends on the conception of the physical according to which things that are physical count as such. For it is according to this conception that only the things that fall under it are duplicated and this is claimed to be sufficient to duplicate everything else. Following this tack, Chomsky challenges those who hold Physicalism to specify what they mean, with an argument that Poland formalizes as following.

P1 There is no relevant *a priori* conception of the physical.

P2 There is no definite *a posteriori* conception of the physical.

C1 Thus, neither an *a priori* nor *a posteriori* approach to assigning content to "physical" is viable.

P3 There are no other approaches to assigning content to "physical."

[41] Originally from Jackson (1994), quoted by Chalmers (1996b, 42).

C2 Thus, there is no definite content assignable to “physical.”

P4 If C2 then C3

C3 There is no definite content assignable to physicalist theses, such as Physicalism.

P5 If C3 then C4

C4 Physicalism is not true and physicalist theses are not substantive hypotheses about the nature of the world.[42]

There are two prime motivations for taking P1 and P2 to be true. One, that folk physics is not a narrow theoretical constraint on our scientific understanding of the world. As Elisabeth Spelke and Kinzler (2007) point out, there is an innate core psychological assumption that the motion of physical objects will be continuous and that space is Euclidean, for example. However, while thinking under assumptions that contradict these principles proves hard, modern science shows that it is not impossible to do so, and our best current scientific theory of space actually requires that we do. And secondly, in extension, the best development of a commonsense *a priori* physics, Cartesian Contact Mechanics, was proved false. So it is simply not true that we can give an *a priori* content to Physicalism.

P2 is just the recognition of the fact that physical theory will change in ways that make any identification of physics with any known physics false. We have good historical reasons to believe current physics will be replaced by another, better physics, an assumption strengthened by the fact that there are fundamental tensions between

[42] Poland (2003, 35). Originally, C3 did not include “…such as Physicalism” and C4 read as “no aspect of the standard metatheory of physicalism is tenable (i.e., physicalist theses are not true, are not empirical hypotheses, play no useful methodological role, and have no human significance). The employed C3 and C4 are just corollaries of Poland’s C3 and C4, respectively.

quantum mechanics and general relativity. This is so even if scientific realism is true, and significant portions of truth will be preserved and augmented in a successor physics. For nevertheless, it will be true that a world described by the physics we now have is a world not identical to ours. So we arrive at Hempel's Dilemma: either our conception of the physical entails the thesis that our current physical theory is completely right, and in that case Physicalism is wrong, or it is the trivial thesis that a future physics will get it right, though we have no idea what that theory is like. Our conception of the physical is open and evolving: it was different for the Cartesians, for the Newtonians, and for modern physicists, so it is hard to say, even for those who try to *assert* Physicalism, just what their commitments are.

Following Chomsky, Poland argues that "Conceptions of the physical are, at best, contingently tied to theories in physics…Since such theories are open and evolving, the concept of the physical is unstable, and hence, not sufficiently well-defined for the purpose of framing empirical and metaphysical hypotheses" (Poland 2003, 33).

P3-P5 and C1-C4 seem uncontroversial to me.

In response to the problem situation, Poland believes that the best conception of the proper commitments of physicalism are methodological rather descriptive. Thus, he has to accept the conclusion that the physicalist thesis "a minimal physical replica of our world is a replica *simpliciter*" is not an *assertible* sentence. For Poland all that can be saved from the descriptive truth-valuable version of Physicalism is the "spirit" of physicalism: a methodological stance regulated by certain ideals about how to solve scientific problems. He proposes:

> On this view, physicalism is understood as embodying a certain "spirit" which involves the deference to science already identified and a commitment to a certain

> sort of program of unification, and it is these commitments which are expressed by physicalist principles and which remain constant as physical theory and our conception of the physical evolves…But methodological physicalism departs from the standard metatheory of physicalism with respect to how physicalist principles are supposed to be construed. The principles are best viewed not as having truth-value, or hence, as being empirical hypotheses. Rather, they function as regulative ideals which both call for, and aid in, the development of theories and associated entities which call for the specified structure (Poland 2003, 38).

Poland proposes to replace Physicalism with a set of instructions for those who try to understand nature. Faced with certain unification problem situations, the regulative ideals call for integration in various forms, and interpreting failure to integrate as a problem in theory, a mystery, or an impossibility (Poland 2003, 41).

Of course, this view is at best just short of capitulation. Physicalism was supposed something that was true or false, and this minimal theoretical desire is thrown out.

It is also unclear how successful its application would be to statements such as "a minimal physical duplicate of our world is a duplicate simpliciter," which are central to philosophical debates, such as the issue of dualism. Would that whole debate have to be recast along non-cognitivist lines?

It also fails to say why such an attitude, "the spirit of physicalism," is theoretically fertile, and if fertile, why should it not be grounded in truth?

2.2.3 A Problem for Micro-Fundamentalism

Philip Pettit argues that a plausible definition of Physicalism just claims that everything empirical is composed of microphysical entities and that macro-entities are composed of those micro-entities, that there are micro-laws, and macro-laws are wholly dependent on and necessitated by those micro-laws (Pettit 1993).

It may look like Pettit's definition gives us a good enough idea of what is meant by "the physical." However, if Physicalism is defined this way then it would be refuted by existing physics. Consider the following point Papineau makes:

> Prepare two electrons in the singlet state and send them off in opposite directions. The left hand electron will have a 50% chance of showing spin-up in the x direction, and 50% chance of showing spin-down. The same is true of the right hand one. They are—let us suppose—a light year apart, and in consequence have no current causal connection. Yet, … the joint state of the two electrons is 'entangled'. If the left hand electron is spin-up, the right hand one will be spin-down, and vice versa. This is a 'non-local' fact about the joint system, in the sense that it cannot be viewed as the sum of local facts about the separated electrons (Papineau 2007).

In the envisaged scenario the particle's behaviour is *not entirely* dependent and determined by some micro-entities. Rather, the entanglement relation between the envisaged particles a light-year apart is actually a macro-entity determining the state of one of the particles; one not wholly determined by or wholly dependent on some other particularly micro set of entities or laws. This is a clear case where microphysicalism does not hold. The example is a particularly cogent case because it comes from physics itself, the holy-land of microphysicalists.

2.2.4 Problems for Papineau's Proposal

Papineau on the other hand proposes that the physical be understood just as what is inorganically identifiable, or "identifiable non-mentally-and-non-biologically" (Papineau 2002, 41; 2007). There are two problems with this proposal. First, mental and biological things exist. But then mental and biological things, by Papineau's analysis, are non-mental non-biological things. Papineau says that physicalism "is to be understood as a matter of property identity" between mental and biological properties, on the one hand, with non-mental non-biological properties, on the other (Papineau

2002, 47). This would seem to be a contradiction, claiming that that which is (as we agree) biological or mental is neither biological nor mental. His project requires that we say that something *is* what it *is not*! We do not negate the existence of the reduced property while asserting the existence of the reducing property when a reduction is found. We do not reduce, say light, by identifying it with something that is not light!

There may be type-type identities between mental properties and "other" properties. Physicalists who think these other properties are uncontroversially non-mental will find that they negate their own project when they posit the identities. This is because by identifying a supposedly uncontroversially non-mental property with a mental property, it becomes, under that assumption, uncontroversially mental. The uncontroversially non-mental property would be identical to a mental property. This project simply cannot be carried out and we know it from the beginning.

There is a second problem with Papineau's proposal. By Papineau's lights, core immaterial entities would be counted as physical. For example, take a kind of immaterial stuff- what some possible ghosts, angels, immaterial minds, and gods are made out of. It is inorganically identifiable. Given, as above, that things are physical if they are inorganically identifiable, this immaterial stuff would be counted as physical. This seems wrong.

2.2.5 Dowell's Alternative Proposal is Not Quite Right

J.L. Dowell proposes that the physical is that which is, or is constitute by, what is "posited by the ideal scientific theory of the world's relatively fundamental elements." (Dowell 2006, 40). Here, the ideal scientific theory is one that has an integrated set of fully well-confirmed hypotheses, and the fundamental elements are those that figure into the unified explanation of the well-confirmed hypotheses (Dowell 2006, 39). Physicalism can then be interpreted as the thesis that a minimal duplicate of the posits of the ideal scientific theory of world's relatively fundamental elements is a duplicate *simpliciter.* She says that this conception satisfies five fundamental constraints on what we can say "the physical" means. One is the Genuine Question Constraint, which says that physicalism is neither trivially true nor trivially false. The second is the Contingency and A Posteriority Constraint, which says that "there are counterfactual worlds in which physicalism is false" (Dowell 2006, 28). The third, Content Constraint, which says that we should be able to "identify what would count as falsifying physicalism both counterfactually and actually…and say something about what an explanatory reduction has to be like to in order to be a physicalist reduction" (Dowell 2006, 29). The fourth is the Explanatory Constraint, which says that non-physical truths are made true by what makes truths about physics true. The fifth is the Conceptual Continuity Constraint, which says that there should be continuity with the pre-theoretical notion of the physical (Dowell 2006, 30-31).

In my view, while the thesis might well be true, there are several shortcomings with the view. One is that it replaces the notion we are trying to understand more, "the physical," with the "fully well-integrated" or "well-confirmed" and "fundamental elements." These terms seem to be subject to the same challenge that Chomsky posed for physicalism. All we have to do is make the appropriate substitutions of "the physical" for "the fully well-integrated," etc. We can say the same thing about these terms that Dowell says about Dowell's posited set of interrelated hypotheses: "Given that we have no idea what will be the posit of that theory, we also have no idea what won't." And we can substitute his "physicists" for "relevant epistemic authority" to denote the makers of Dowell's ideal theory, and her "ultimately developed" for "ideally developed" thus: "And given that we have no idea what won't be a posit of the theory *ideally developed* by the *relevant epistemic authority,* we're unable to identify what would count as falsifying physicalism on the resulting formulation."[43]

I agree in principle however that ideal physical theory would be well-integrated, well-confirmed, and about fundamental elements. However, I fail to be much informed in quite the necessary way in the present context. For example, in a world where Descartes' Substance Dualism is true, its systematicity would make it well-integrated, its truth would provide the grounds for it to be well-confirmed, and it would explain by virtue of fundamental elements, including positing immaterial minds. By Dowell's definition, if we find ourselves in such a world, Physicalism would be true. This seems wrong. This violates the Genuine Question Constraint and the Content Constraint in its failure to count Physicalism as false in those worlds, supposing them to be actual for the counterfactual reasoning she invites us to engage in.

Of course, *given that our world is not such a world*, Dowell's definition would be true. The thing to notice is that it would fail to tell us the opposite if we lived in a Substance Dualist world. For this reason, Dowell's theory fails in a significant way to satisfy his Genuine Question Constraint and the Content Constraint.

At one point Dowell might seem to tap this hole when she says that "to count as basic and physical, a property must be well-integrated into the most complete and unified explanation possible for the relatively most basic occupants of space-time." It seems that it does because it tempts us to suppose that what is available for explanation,

[43] Dowell (2006, 37), italics added to signal the substitutions.

according to such a theory is all in space-time. But in fact it is not ruled out that relatively fundamental mental elements outside of space time contribute in an explanation of the occupants of space-time. Thus, in a world where Substance Dualism is true, the most complete and unified explanation possible for the relatively most basic occupants of space-time will include reference to immaterial mental entities. Since, for Dowell, "we define 'physical theory' as a scientific theory of the world's relatively fundamental elements," (Dowell 2006, 39) and in Descartes' theory, immaterial minds are relatively fundamental elements, then, were Substance Dualism to be true, Dowell's theory would count immaterial minds as physical, and therefore figuring in the total set of explanations to be generated from theory, including explanations of the occupants of space-time.

Dowell believes we think this is implausible only because we think it is very unlikely that a final theory of the kind she defines will posit relatively fundamental minds outside of space-time (Dowell 2006, fn. 28). In that case maybe she could have saved us the trouble of trying to make a significant theory and said: Whatever the world turns out to be like, let T the ideal theory that describes it. Let Physicalism be the thesis that T is true actually. Therefore Physicalism is true. Of course, by this definition, Physicalism is true, no matter what, but trivially and without the content required of it by the dialectical context.[44]

2.2.6 A Modest Proposal that Get Us What We Want

I think we can maintain not only the methodological attitude Poland mentions, but in accordance with the empirical status most physicalists ascribe to their theses, the

[44] Compare Wilson 2006, where she makes comments in the same family as mine. My distinctive emphasis is in showing how Dowell's proposal fails her own stated theoretical goals.

belief that the ideal of a unified conception of nature is realized, or made true, by nature. That is, the methodological aspect of Physicalism is good precisely because nature itself is unified in the ideal way and our unifying "spirit" is satisfied by nature. Interpreted this way Physicalism attains the status of making more traction with nature since it can be confirmed and disconfirmed as we see science developing, and so the issue of physicalism, and particularly Physicalism, becomes empirically decidable.

Some things physicalists should, intuitively, claim to be true are:

1. The physical world is lawfully regulated.[45]
2. Physical entities exhibit themselves in space-time.
3. Neurons are physical entities.
4. Being a moving of a square is a physical property.
5. The stop-light's changing from green to red is a physical event.
6. Entangled systems are physical.
7. Angels, gods, ghosts, are not physical, if they exist.
8. Platonic entities are not physical, if they exist.

I think conceptions of the physical from such diverse and central thinkers as René Descartes and Donald Davidson would accept these statements (had Descartes known of neurons, entangled systems, and stop-lights). Notice further, (3) to (8) can be explained in terms of (1) and (2). For we might think that it is in virtue of the fact that neurons, moving squares, and stop-lights' changing colour, are lawfully regulated and exhibited in space-time, that they are physical, and for the physicalist all exhibited properties, events, and objects are in a quite significant sense physical. Angels, gods, and ghosts, on the other hand, are typically conceived as having powers that can

[45] Laws can be probabilistic.

override the laws of nature, and at least sometimes, to exist outside space-time. If putative angels, gods, and ghosts lose these powers, I see little reason to call them by such names. Platonic entities are typically also conceived as being outside space-time and the causal flux of nature, and not regulated by its laws.

A hypothesis naturally arises: To say then that all is physical is to say all is wholly regulated by the laws of nature and wholly embedded in space-time.

Physicalism is the contention that everything that exists has these qualities. This results in:

> **Unpacking Physicalism:** A minimal duplicate of the lawfully regulated existents within space-time is a duplicate *simpliciter* of our world.

This conception rejects P1 and P2 in Chomsky's Challenge. Our conception of the physical has an *a priori* component (in the sense that it is constant in its historical and philosophical application-condition)- the core description: *that which is lawful and wholly in space-time*. And P2 is also false since we know plenty of things that are lawful and in space-time, and that as we continue our epistemic journey into the physical, if all goes well, the truth-value of our beliefs tend towards being completely true. Therefore, "a physical duplicate of our world is a duplicate *simpliciter*" is an assertible sentence.

Further, notice that this proposal does satisfy Dowell's Constraints. First, there is genuine question about whether all is in space-time and regulated by the laws of nature. It is not trivially evident that this is the case. Secondly, there are counterfactual worlds in which not all occupants of the world are in space-time and regulated by the laws of nature. The third, we are able to identify what would count as falsifying physicalism both counterfactually and actually, for it is possible that not everything is in space-time or is regulated by the laws of nature.

I am not sure what informative value reductionists will find in any “physicalist reduction” but what my theory offers is just that looking for the ways in which the occupants of space-time are lawfully integrated will be as informative as such reductions get. Thus, light’s being physicalistically reduced to electromagnetic radiation reveals something about the nature of light and its interactive potential. Fourthly, were non-physical truths to have truth-makers, they would be made true by physical facts. Fifthly, *Unpacking Physicalism* exhibits an obvious continuity with the pre-theoretical notion of the physical, and historical continuity with the theoretical alliances that define the physicalism debate- including providing a contrast to Substance Dualism (Dowell 2006, 30-31).

Further, this account is not self-defeating, does not count immaterial things as physical, and is not contrary to current physics. For this reason, I think it is a conception worthy of the content of Physicalism.

It should be noted that this conception implies that the mental is physical, given that mental properties or their instances are in space-time and governed by the laws of nature. Since physicalism is a monism, it should come as no surprise. It would be a refutation of physicalism if the mental was not physical. Some may have some resistance to this idea because there is a categorical difference between the physical and the mental. Indeed there sure are important differences between fearing a lion and having a spin up, or perceiving a fox and floating down the Amazon River. But likewise, there are important differences between having a spin up and floating down the Amazon River. The conception proposed here allows for this important kind of diversity, while explaining what is importantly similar about all these things, including those things that are mental, in virtue of which they are all physical.

2.2.7 The Mental Comes Back to Power

Now that we have elucidated what it is to be physical, and otherwise independently motivated such a conception as the one to be held when one holds Physicalist theses, it will be useful to look at how Kim's Problem of Causal Exclusion is affected. Let us look into it bit by bit. Suppose, again, we have an instance of mental to mental causation. This would be a situation represented like this:

1. An instance of a mental property M1, occurring at t1, causes an instance of the mental property M2 occurring at t2.

If we are to be Physicalists, then there must be an important sense in which M1 and M2 are not only mental properties, but also physical properties. Since Physicalism is the doctrine that everything that exists is in space-time, and is subject to lawful regulation, if mental properties exist, they must be physical. One way in which they are is that instances of mental properties are physical. While this works with trope conceptions of properties, some may want to hold a Platonist view of properties, and for them *properties as such* can never be physical. Since Platonists conceive of properties as outside of space-time and the causal flux of the universe, properties for them could never enter into causal interactions, whether physical or mental. However, they should allow that there is significant sense in which at least some properties, like having particular mass, are physical, and that having such properties enables the things that have them to enter into causal interactions of particular sorts. One way to do this is to say the following:

Physical Sufficiency Condition on Properties: properties are physical if their instances are physical.

Physical Sufficiency Condition on Instances: instances of properties are physical if they are in space-time and regulated by the laws of nature.

Causal Sufficiency Condition on Properties: properties are causal (or have causal powers) if their instances cause (or have the ability to cause).

As applied to the particular case in question we might note that M1 and M2 can be said to be physical properties, if their instances are physical. Given Physicalism, instances of M1 and M2 must be physical, as well as mental. Given the conception of the physical argued for before, the M1 and M2 instances are wholly within space-time and regulated by the laws of nature.

Kim says that it follows from our analysis of supervenience that:

2. M2 has a physical supervenience base P2, at t2.

But it seems unnecessary to stress that M2 must have a physical supervenience base, P2. M2 is already physical, so there seems to be no reason to say that M2 needs some physical supervenience base, apart from its own existence, to exist. Given that what makes M1 and M2 physical is just the general fact, that like with any other physical property, the instances are physical, the question Kim wants to subsequently ask, “What made M2 come about on this occasion?” is not independently compelling. This is because there is no apparent problem about M2’s integration into the causal network of nature that does not apply to P2 as well- they are both, by Physicalist standards, equally physical and therefore equally within space-time and regulated by the laws of nature. Nevertheless, Kim says we get these possible answers to the question:

3. M2 is instantiated on this occasion: (a) because an instance of M1 caused an instance of M2, or (b) because an instance of P2, the M2-instance's physical supervenience base, is instantiated.

Kim says that the problem is that we have two competing explanations for the occurrence of this instance of M2 and P2 seems like an attractive candidate. Here's why: The occurrence of M2 is necessarily co-occurrent with one of its realizers or bases. At least in the circumstances, P2-instance is sufficient for the M2-instance. But of course, there are different "why" questions, one may note. If you ask "why?" in search of simultaneously occurring instantiation of properties that subserve the thing you wish to explain, then (b) is attractive. On the other hand, (a) picks out something (usually) located at another place in space-time, one other than where the cause is. As such there is no conflict between the two different answers to the *two different questions*. Suppose someone asks you "why is this atom a Carbon atom?" You might go and say that the atom in question has a total of six protons and six electrons, and that anything that has an atomic structure with a total of six protons and six electrons is a Carbon atom, as Kim recommends in saying that (b) is an appropriate answer. This is indeed a perfectly legitimate answer to the apparently relevant question. But suppose that this is not what the person asking wants to know, and she clarifies that what she wants to know is about the causal ancestry of the atom in question. The person asking wants to know something else, some other facts about what the Carbon's existence depended on.

Suppose that the Carbon in question was formed through fusion in a star core. In that case, the appropriate answer is more like (a). In that case, we say that the Carbon atom was formed in stellar fusion, and that this is a process in giant and super-giant stars that weigh down on their own cores with sufficient force to generate the triple collisions between Helium nuclei that in turn activate the strong force, which binds them together into Carbon nuclei, which, through the electromagnetic force, go on to acquire six relatively slow-moving electrons. This is also a legitimate answer as to the question of "why is this atom a Carbon atom?" This answer is like (a). If someone says that this second explanation is somehow illegitimate, because it conflicts with the reductive explanation given before, they are certainly wrong. There is no conflict as long as we understand the reasonable assumption that different questions many times have different answers. Kim urges, nevertheless, that we must somehow try to vindicate the instance of M1's causing the instance of M2. This is unnecessary as can be seen from the fact that M1 and M2 are as physical as their supposed subvenient bases, and that different "why" questions command different answers, without an implied conflict.

4. The M1-instance causes the M2-instance by causing P2.

Thus, 4. asserts that a supervenient mental property instance causes a subvenient physical property instance. This shows that: "Under the mind-body supervenience assumption, mental-to-mental causation implies, or presupposes, mental-to-physical causation" (Kim 1998, 43). So, the question now becomes whether we can make sense of mental to physical causation. From what we have discovered about the physical, there is no mystery about mental-to-physical causation, since the "mental" in "mental-to-physical causation" is itself physical- as much the "physical" within it is. As such, the interaction can also be described as “physical-to-physical,” and this is not a kind of causation at risk here. Kim continues:

5. The M1-instance itself has a physical supervenience base, the P1-instance.

Now it looks like the P1-instance is going to preempt the M1-instance from causing the P2-instance. Since given Physicalism, every physical instance that has a cause has a physical cause, the P2-instance, being a physical instance, has a physical cause- the P1-instance being the relevant candidate. What further causal work is there left for the M1-instance to do? It seems like the M1-instance as a cause of the P2-instance, and therefore of the M2-instance, is dispensable, Kim thinks. However, it is clear that the M1-instance is physical, so there is no problem about how the M1 could cause the P2. For the conflict was supposed to surface when we consider that every physical effect that has a cause, must have a sufficient physical cause, what is called the Principle of the Causal Closure of the Physical. But in this case we do not violate this principle, since M1 is not only a mental property, it is a physical property, since it satisfies the Physical Sufficiency Condition on Properties.

The resulting alternative Kim contended is:

6. The physical P1-instance causes the physical P2-instance and the mental M1-instance supervenes on the P1-instance and the mental M2-instance supervenes on the P2-instance.

This yields the result, Kim believes, that if the properties instantiated at a subvenient level are distinct from their resulting properties at a supervenient level, these latter properties are causally inoperative. What does all the work is exhaustively the subvenient base. These constitute the real causal powers of any instance of apparent supervenient causes, he thinks. However, our conception of the physical has had the effect of truncating each step Kim wants us to follow him along. Under our conception of the physical, there is no reason to believe the mental is some sort of nomological dangler, but is rather understood quite smoothly as existentially and causally on a par with any other thing in nature. To use a frequent example in the literature, suppose that being an instance of having a pain is admitted to have a supervenience base of being an instance of a c-fiber firing. We might ask why being an instance of c-fiber firing is admitted to be physical but not having a pain. If the answer is that having pains are mental, we argue that there is a double standard implicit here, characteristic of dualism. Why hold the mental to not be physical, but the neurophysiological, as c-fibers firing are, to be physical? There is no good reason. Both have as good a claim to being physical, and if you are a Physicalist, there should be no question that mental things are physical in the way required for everything else in the universe to be physical. Further, if the conception of the physical argued for here does indeed help dissolve the Problem of Causal Exclusion, then this is an argument for that conception as well. However, there may still be some force in Kim's argument, if the notion of supervenience is a significant generator of the threat to the mental in satisfying the Causal Sufficiency Condition on Properties. One way of beginning to see this is to revise 3 to 3'.

3' M2 is instantiated on this occasion: (a) because an instance of M1 caused an instance of M2 (as an instance of supervenient causation), or (b) because an instance of P1, the M1's instance's physical supervenience base, causes M2 to be instantiated (perhaps by causing P2, M2's physical supervenience base).

Formulating the situation this way, the idea that there is competition between the instance of M1 and the instance of M2 is not ruled out so easily. We will tackle this version in chapter 6.

2.3 Kim on Non-Basic Causation

Jaegwon Kim is the originator of the Problem of Causal Exclusion and it is therefore instructive to see what his historical development of this problem and his solutions are in order to see what is at stake and what a successful solution could be. For the moment, we will critically look at Kim's development of the Causal Exclusion Argument, without the help of the clarification of Physicalism obtained in the last section, so that our exposition more closely resembles Kim's.

2.3.1 Supervenience and Dependence

The immediately following sections in this chapter will assume that the relation of supervenience is a relation between basic and non-basic properties of things, where the superveniently non-basic properties of things are determined by the subveniently basic properties of things. I believe that the dichotomy of the basic versus non-basic captures what Kim means when he speaks of supervenience, i.e. the dichotomy between the subvenient versus supervenient. The next sections will discuss Kim's views on some relevant theses and will assume subvenient and supervenient properties of objects just are basic and non-basic properties of objects respectively.

I think this assumption about the supervenience relation is well-grounded in the tradition from which it originates. The notion of supervenience is a philosophical notion *par excellence*. Unlike the notion of cause, its sense in philosophical contexts is much to be determined by the philosophical work to be done by it, while commonsense intuition is not. Kim and McLaughlin note that it was Davidson who introduced the notion of supervenience into the centre-stage of debates in the philosophy of mind. Davidson did this with the following passage:

> Mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken to mean that there cannot be two events exactly alike in all physical respects but differing in some mental respects, or that an object cannot alter in some mental respects without altering in some physical respects (Davidson 1970).

Mental characteristics, as exhibited in our world, might well be physical. But what is assumed in much of the discourse about how the mental supervenes upon the physical is that mental properties are dependent on (perhaps other) physical subvenient properties, and are properties that are not wholly in themselves mental. Thus, what is assumed is more generally this: “First, supervenience is to be a relation of dependence: that which is supervenient is dependent on that on which it supervenes” (Kim 1989a, 139), where “supervenient properties are dependent, or are determined by, their *base* properties” (Kim 1989a, 140; emphasis added).[46] “Secondly, supervenient dependency is not to entail the reducibility of the supervenient to its subvenient *base*” (Kim 1989a, 139; emphasis added). Thus, mental properties are taken to be supervenient properties dependent on, or determined by, base properties. Mental properties, like all supervenient properties are not basic; they are determined by base properties. It is typical for writers on supervenience and its applications to call subvenient properties “base properties” or “basal properties.” It is easy to construe this terminology as an acknowledgement that subvenient properties are basic properties of things with respect to supervenient properties of things, and that since supervenient properties of things are dependent on basic properties of things, but not reducible to them, that they are in this sense non-basic.

For properties, “B-properties supervene logically on A-properties if no two logically possible situations are identical with respect to their A-properties but distinct

[46] He calls this the principle of Dependence.

with respect to their B-properties "(Chalmers 1996b, 35). Logically supervenient properties are such that they cannot vary, given the status of their supervenience base. From an informational point of view, if you have the adequate definitions of supervenient properties, you can in principle read them off their supervenience bases.

The other important class of supervenient properties is of those that belong to natural supervenience, which according to Chalmers, hold where the laws of nature hold: "B-properties supervene naturally on A-properties if any two naturally possible situations with the same A-properties have the same B-properties" (Chalmers 1996b, 35). As an example of natural supervenience, Chalmers points to:

> the pressure exerted by a mole of a gas systematically depends on its temperature and volume according to the law pV=KT, where K is a constant…In the actual world, wherever there is a mole of gas at a given temperature and volume, its pressure will be determined: it is empirically impossible that two distinct moles of gas could have the same temperature and volume, but different pressure…But this supervenience is weaker than logical supervenience. It is logically possible that a mole of gas with a given temperature and volume might have a different pressure; imagine a world in which the gas constant K is larger or smaller (Chalmers 1996b, 36).

One may be initially unconvinced that a case has been made for the distinctive existence of natural supervenience. For one, it needs to be argued that by changing the K constant, you have not thereby changed the supervenience base, just as you would change it if you changed the temperature, and so would expect pressure to vary. For if the temperature value, mole existence, and K constant are held to be the A-properties, the pressure, or B-properties, supervene logically and natural supervenience collapses into logical supervenience.

Perhaps there is a distinction to be made in terms of whether the laws of nature are allowed to be supervenience base properties, however. Thus we would say that naturally possible situations are those in which the laws of nature are held constant, while other properties in accordance with the analysis above. However, as it stands the analysis of natural supervenience counts cases of logical supervenience as cases of natural supervenience since wherever it is the case that any two naturally possible situations with the same A-properties have the same B-properties, these will be logically possible situations with identical B-properties. For instance, being red supervenes logically on being scarlet red, since any logically possible situation with the property of being scarlet red will be a situation with the property of being red. However, it is also the case that any naturally possible situation with the property of being scarlet red will be a case with the property of being red. By the definition of natural supervenience above, such supervenience would be natural. But natural supervenience cases are not supposed to be cases of logical supervenience, nor vice versa. This is why Chalmers distinguishes natural supervenience from the logical variety, by holding that the former could be made to vary by changing natural laws. Logical supervenience holds between properties, no matter what the natural laws are. The case of red supervening on scarlet red is not supposed to be a case of natural supervenience, for scarlet red things are red no matter what natural laws are in place (consistent with being scarlet red in the first place). This is supposed to be a case of logical supervenience!

We need an analysis that does not count cases of naturally possible situations, where nevertheless logical supervenience holds, as cases of natural supervenience, and vice versa. In order to make this necessary distinction, the analysis looks like this: "B-properties supervene naturally on A-properties if any two naturally possible situations with the same A-properties have the same B-properties, and there is a naturally possible situation and a logically possible situation that have the same A-properties, but the two situations do not have the same B-properties." Thus, in our world pV=KT holds. If you set the supervenience base consisting of one mole of gas at 200 Kelvin with a volume of 200 cubic centimeters you set the pressure it exerts at a value- its supervenient property. However, there is a possible world where things are different. In that world pV=2KT holds, and in that world the pressure exerted will be double its actual value even though you have the same supervenience base properties.

With the view that we wish to understand Kim's argument in terms of this analysis, and Kim's argument, as I presented it, is formed in terms of instances of properties, let us analyze logical supervenience in terms of instances of properties. Kim's argument can also be reconstructed to be framed in terms of properties, and then the above analysis will be the useful one.

For property-instances, "Instance of property B supervenes logically on instance of property A if no two logically possible situations are identical with respect to their A-instance but distinct with respect to their B-instance." The intended interpretation of this analysis is that just like properties can be the same across possible situations, so can instances of properties. There are several contending analyses for such cases, and the reader can interpret this according to her general approach. No new problems will arise that are not already present in the interpretation of sameness claims about properties and individuals across possible situations.

2.3.2 A Fundamental Tension

Now, to the main part of the chapter. A common thread in Kim's work is that he endorses three theses.

A. Real Non-Basic Existence (*Real*): That things with non-basic properties exist, with their own causal powers.

B. Non-Basic Existence is Epiphenomenal (*Epiphenomenal*): But that since irreducibly non-basic properties of things are wholly causally dependant on basic properties of things, irreducibly non-basic properties of things are causally epiphenomenal with respect to basic properties of things.

But, we are not justified in being committed to completely epiphenomenal properties of things as real (by Alexander's dictum), and therefore non-basic properties of things do not exist. However, by *Real*, they must exist, and the way to vindicate them is to say that:

C. Identical Existence (*Reduction*): Non-basic properties of things are identical with basic properties of things that have them.

The tension arises because a property of a thing is either basic or non-basic, but not both. You cannot say that you are ontologically committed to non-basic properties just by being ontologically committed to basic properties! I will argue here that Kim never takes a successful stand on these issues. Although he changed his **explicit** view from believing that mental causation, in which a non-basic property of certain beings purportedly plays a causal role, was a particular form of causation (epiphenomenal causation), to the belief that such "causation" was fake- or not real- causation, he constantly oscillates between these two poles in formulating his own views. On the one hand, he wants to be committed to the keeping of the non-basic property as reduced and vindicated thereby, and on the other hand, dispensing with it as causally inefficacious and therefore unreal. His attempt to reconcile the two views is to identify the epiphenomenal property with the causal and basic property.

The Thesis and the Task

I believe that the explanation of the continuity in Kim's problem with taking a stable stand on the metaphysics of mental causation is that in the last analysis his views on the causal powers of non-basic features of the world have remained fundamentally the same. What has changed is the recognition that epiphenomenal causation is not causation at all. Accordingly, my method of showing this is to look for the historical places in his work in which he (i) Holds that epiphenomenal causation is real, (ii) Changes his views on whether epiphenomenal causation is real, and (iii) The way Kim regards non-basic causation (of which mental causation is an example) now. Having identified the critical historical junctures of Kim's thought about epiphenomenal causation, we shall see where he stands with regards to theses *Real*, *Epiphenomenal*, and *Reduction* at each juncture, and show their tension with one another.

Before executing the task just described, however, there are some terminological clarifications to be made. It is sometimes said that properties have causal powers. Platonists about properties can consistently say such things only if their commitment in saying this is that the things that have them have causal powers, or that the properties' instances have causal powers. That is, those properties *had by a thing* that has causal powers enable it to enter into causal relations with the things they have a power to cause. Platonic properties themselves do not enter into causal relations since they are theorized to be outside the causal flux of nature, and therefore cannot straightforwardly, without further analysis, be said to have causal powers. By the same token, when we say that mental properties cause such and such putative effect-properties, we mean that the thing that has mental properties was enabled by those mental properties to cause such and such putative effects. In what follows it is generally assumed that what is related in causation are events and the locution that "properties of events" means "properties either had by the event in question or by the object of the event in question." At least for the time being, we can regard this as a *prima facie* plausible crutch. It is left open whether the things that have properties that enable them to enter into causal relations are events or substances. It is enough for the discussion to take place to know that "properties of events" expresses the character of events that enables them to enter into causal relations.

Further, Kim seems to flick back and forth in talking about properties and events when he talks about supervenience. However, much of this can be consistently made sense of if we assume that when we say that properties supervene on others, we mean properties of the things that have them supervene on other properties of the things have them. Thus, if we say that mental properties supervene on physical properties we mean that the mental properties of the thing that has them supervene on the physical properties of the thing that has them.

2.3.3 Epiphenomenal Causation as Real and as Hoax: Through the History

In his "Supervenient and Epiphenomenal Causation" of 1984, Kim argues that there is a particular variety of causation that pertains to higher-level, or supervenient, properties. Higher-level, or supervenient properties are non-basic properties. They are properties of things determined by lower-level subvenient properties of the things that have them, basic properties in virtue of which the supervenient properties occur. Kim argues that what does the causal work in such cases are just the lower-level subvenient phenomena, and the “horizontal” relation that exists between higher-level causal relata and lower-level causal relata is that of reducibility. Higher-level “causal relations, whether only apparent or real, are reducible to more fundamental causal relations” (Kim 1984, 94). Higher-level causation, understood in this way, Kim argues is the referent of “epiphenomenal causation.” In other words, Kim argues that the causal connections between non-basic phenomena are epiphenomenal causal connections, reducible to causal connections between basic phenomena.

In 1993 he continued to hold the view that epiphenomenal causation as depicted in his 1984 article was a case of genuine causation. He writes in his “Mechanism, Purpose, and Explanatory Exclusion,” for C as a mental cause, that “we must think of the causal efficacy of C in bringing about E as dependent on that of its physical correlate C*” and comments on this statement: “I believe this picture can be generalized; see my ‘Epiphenomenal and Supervenient Causation’” (Kim 1989c, 247). However, he also says that in cases such as this one, there are not “two explanations for one explanandum; for epiphenomena do not explain.” (Kim 1989c, 244). We might think that the apparent contradiction is only apparent since Kim holds that the superveniently epiphenomenal cause is reduced to its subveniently causal base. In that case the supervenient cause is identical to its subvenient base. However, the contradiction re-arises when we think that an *epiphenomenon,* at the higher-level, is just the *non-epiphenomenon* that grounds it. Further, Kim implies by the above quote, that since it is a necessary condition on causal explanation that the properties alluded to in the explanans have causal powers, that epiphenomena do not causally explain. This is a reasonable idea, but then he is implying that the referent of the explanans both is and is not an epiphenomenon. Restricting the scope of views to be considered at this point to the views he articulates in the *Supervenience and Mind* collection, it is only in 1993 that he articulates more explicitly and sharply his new view that epiphenomenal causation is not real causation at all.

In his "The Non-Reductivist's Troubles with Mental Causation" of that year, Kim articulates the second part the Problem of Causal Exclusion. According to this part, it is necessary for there to be mental to mental causation that there be mental to physical causation. That is, given that at the time the mental effect occurs, it is sufficient that its underlying physical realization base occur for it to also occur, the putative mental cause must cause the physical realization base of its mental effect. However, if the physical and mental causes are different, the physical cause preempts the mental cause from doing its putative causal work. This sort of preemption is justified by the idea that the physical world is causally closed under physical law. The resulting picture is that the putative mental cause is "treated as an epiphenomenal dangler from its physical realization base, with no causal work of its own to do" (Kim 1993a, 355).

Kim holds this view about putative epiphenomenal causation to this day, as seen in his major works *Mind in a Physical World* from 1998 and the subsequent symposium on it, and *Physicalism, or Something Near Enough* from 2005, and his work in between as seen in his "Blocking Causal Drainage and Other Maintenance Chores with Mental Causation." In his 1998 study he considers program explanation, according to which higher-level properties explain because they "program for" the having of some property or other that does some work in bringing about the effect in terms of which the higher-level properties were defined. Thus, in the non-mental realm, the fragility of the base explains its breaking by a light bump because “fragility” programs for some molecular property that made it so that a light bump breaks the base. In the mental case, being hungry would explain my going to the kitchen if I know food is found there because “being hungry” programs for some molecular configuration that makes me go to the kitchen if I know food is found there. Kim (1998, 74) considers such properties as fitting the model of supervenient causation, a variety of epiphenomenal causation, one he advocated in his 1984 article. Of these properties he says that “it is clear that program explanations, whatever their explanatory value, cannot be causal explanations” (Kim 1998, 75). We may immediately infer that epiphenomenal properties, those alluded to in higher-level non-basic explanations, are not considered to be causal. This is because, according to Kim, it is never in virtue of having these properties that the thing that has them causes an effect to be explained, and therefore causal explanations that make reference to those properties are not successful causal explanations. In his 2003 article he alludes to the possibility that there is a causal connection between supervenient higher-level (non-basic) phenomena as he advocated in his 1984 article. Diagrammatically, that connection can be expressed as a horizontal arrow connecting the two representations of the events. But he says of this connection/arrow: “inserting a broken-dot arrow and calling it “supervenient” causation, or anything else (how about “pretend” or “faux” “causation”?), does not alter the situation one bit. It neither adds any new facts nor reveals any hitherto unnoticed relationships. Inserting the extra arrow is not only pointless; it could also be philosophically pernicious if it should mislead us into thinking that we have thereby conferred on M, the mental event, some real causal role" (Kim 2003, 171). This view about epiphenomenal causation is sustained in his 2005 book, where he remarks that in case the mental fails to reduce to the physical "we are then faced with the specter of epiphenomenalism and we must learn to live with causally impotent mental properties" (Kim 2005, 159).

In sum, there are three relevant stages in how Kim regards epiphenomenal causation. In 1984 such causation was real and pervasive. He maintains this view through to 1989, with some very apparent tensions, and by 1993 he articulates the Problem of Causal Exclusion, which shows that epiphenomenal properties are excluded from entering into causal transactions, and he disavows epiphenomenal causation as real causation. The view expressed there has been constant since.

2.3.4 Kim's Views on *Real, Epiphenomenal,* and *Reduction*: Through the History

1984: Kim Holds *Real, Epiphenomenal, and Reduction*

In his 1984 article he believes *Real:* That things with non-basic properties exist, with their own causal powers. Under the assumption that causal relations between events entail laws connecting properties of the causing event and the effect, when Kim asserts that "central cases of epiphenomenal causation will be seen to involve "real" causal relations and that epiphenomenal causal relations of this kind are pervasively present" (Kim 1984, 94), he is committed to the idea that such properties of the "central cases" of epiphenomenal causes are real. Further, he considers the relata of the central cases of epiphenomenal causation to be mental and macro events. Under the assumption that mental and macro properties of things, including events, are not basic, but are rather supervenient on the basic, it follows that things with non-basic properties exist, with their own causal powers. That is, in 1984 Kim endorses *Real*.

At this time he also endorses *Epiphenomenal:* That since irreducibly non-basic properties of things are wholly causally dependant on basic properties of things, irreducibly non-basic properties of things are causally epiphenomenal with respect to basic properties of things. Kim begins his discussion by citing the occasionalist Jonathan Edwards (1758), and Wesley Salmon's illustration of a pseudo-process. According to Edwards, each temporal slice of a material object is created *ex nihilo* by the will of God. The result is that there are apparent persisting material objects. But the fact of the matter is that they are like images on a mirror, "which seem to remain the same, with a continuing perfect identity… [though] it is known to be otherwise…so that the image impressed by the former rays is constantly vanishing, and a new image is impressed by new rays every moment." According to Salmon, a pseudo-process is an apparent causal process the constituents of which are not causally related. An example is the apparent causal process of a spot of light moving on the wall. It may appear that the spot is moving itself about, or that its position an instant later is causally determined by its position just an instant before, but we know this is not the case. Kim believes that the properties of non-basic things are, like Edwards' material objects and Salmon's pseudo-processes, wholly dependent at any one time on basic properties of the things in question. He says that what is common between Edwards' material objects and Salmon's pseudo-processes, and epiphenomenal causation is that "they all involve at least *apparent* causal relations that are *grounded* in some underlying causal process" (Kim 1984, 94). For Kim, the underlying causal processes are micro-processes, upon which all else depends, or is grounded on. He endorses the doctrine of micro-determinism, according to which the macro-world is the way it is because the micro-world is the way it is. More formally, “worlds that are microphysically identical are one world...” (Kim 1984, 100).[47] For Kim, microphysics is what is basic. Once the microphysical facts are arranged, the rest of the facts are arranged. It follows that all that is not microphysical is determined by the microphysical, both causally and superveniently. It follows that at any one time, the non-microphysical, or non-basic, properties of things are wholly determined by microphysical properties of things. Now suppose that non-basic properties of things have causal powers, as Kim does in endorsing *Real.* We are now confronted with the question of, what causal work instances of them do in addition to the causal work done by instances of the microphysical properties of microphysical processes upon which such powers of non-basic properties depend. By his doctrine that were powers of non-basic properties (distinct from the basic properties of things) to have causal influence on subsequent events they would violate the closed character of the physical (Kim 1984, 96), it follows that non-basic properties do no causal work. In this traditional sense, non-basic properties of things that have them are epiphenomenal with respect to the basic properties in virtue of which such non-basic properties are had by anything that has them. It follows that Kim endorses *Epiphenomenal.*

Now Kim also endorses *Reduction* in 1984. Non-basic properties of things are identical with basic properties of things that have them. Kim believes all non-basic properties of things, things “pervasively present all around us,” are reducible, through

[47] There may be a distinction between the micro –for example if there are micro-ideas- and the microphysical. However, it is clear that for Kim micro-determinism is microphysical determinism.

proper scientific investigation, to basic micro-phenomena. To support this Kim argues that the scientific research strategy is “to try to understand the behavior and properties of objects and processes in terms of the properties and relationships of characterizing their micro-constituents.” The ground of this strategy, Kim thinks, is the “belief that macro-properties are determined by, or supervenient upon, micro-properties…In this global microdeterministic picture there is no place for irreducible macrocausal relations” (Kim 1984, 96-97). Thus, reduction, which exhibits an identity between a phenomenon we understand in different ways is a way to legitimize ontological commitments towards non-basic phenomena. When a reduction is successfully completed, the diverse ways of epistemically accessing the phenomena are said to be accessing *the same* phenomena, referring to the same objective properties and other entities of the world. Accessing or referring to the phenomena is legitimate (or successful) because it resolves itself into accessing or referring to basic phenomena as *vindicated* by microphysics. In application of this fundamentalist reductionist idea, “Causal relations that resist microreduction must be considered “causal danglers,” which like the notorious “nomological danglers,” are an acute embarrassment to the physicalist view of the world” (Kim 1984, 100). Note that Kim is talking about causal relations, not explicitly properties or events. However, since things with no causal powers are not admitted into our ontology, non-basic features of the world, like macro and mental properties, events, relations, and objects must be vindicated, the causal processes they enter into must be “reducible to more fundamental causal relations” (Kim 1984, 94), “in general reducible to microcausal relations” involving microreductions of the causal relata (Kim 1984, 99). Thus, in 1984 Kim endorses *Reduction* as well as *Real* and *Epiphenomenal.*

1993: Kim endorses *Real, Epiphenomenal, and Reduction*

In 1993 he holds *Real* as well, even though, by this time, he considers epiphenomenal causation as being implausible as a variety of causation. Here, where Kim is powerfully wielding the Problem of Causal Exclusion, when he considers what counts as physical he is lax with everything but the mental. He follows Hellman and Thompson in saying that physical properties are those referenced in current theoretical physics, and he says that their strategy can be extended to higher-level properties referenced in chemistry and biology. Mental properties of things, he says (Kim 1993a, 340) depend on "physical-biological" properties.[48] Since higher-level chemical and biological properties are non-basic properties of things, it follows that things with non-basic properties exist. Since Kim (Kim 1993a, 348) accepts Alexander's dictum, that "to be real is to have causal powers," and just a year before, he held the congenial Causal Individuation of Kinds Principle, according to which "kinds in science are individuated on the basis of causal powers" (Kim 1993c, 326) Kim is committed to recognizing things with chemical and biological properties, which are non-basic properties from the perspective fundamental physics. Thus, in 1993 Kim is committed to *Real.*

At this time Kim is also committed to *Epiphenomenal.* Kim articulates the most lucid form of the Causal Exclusion Problem to that date in his 1993 "The Non-reductivist's Troubles with Mental Causation." In it, he says: "So suppose M is causally efficacious with respect to some mental property M*, and in particular that

[48] It is unclear why there can be "physical-biological properties" but not "physical-mental" properties.

some instance of M causes an instance of M*. But M*, *qua* mental property, is physically realized; let P* be its physical realization base. Now we seem to have two distinct and independent answers to the question: Why is this instance of M* present? *Ex hypothesi*, M* is there because an instance of M caused it; that's why it is there. But there is another answer: it is there because P* physically realizes M*, and P* is instantiated" (Kim 1993a, 351). Barring the options of systematic independent explanations for mental instances, Kim argues that the only way in which we may try to vindicate the denial of the thesis that M is realized by P, and M* is realized by P*, while M does no causal work in bringing about the instance of M*, is to have downward causation, whereby M instances cause M* instances by causing P* instances, upon which M* instances supervene. We may infer the following general principle:

> **The Causal Realization Principle:** If a given instance of S occurs by being realized by Q, then any cause of this instance of S must be a cause of this instance of Q (and of course any cause of this instance of Q is a cause of this instance of S) (Kim 1993a, 352).

However, Kim thinks that the only potentially open door for the non-reductionist is not a real alternative. Downward causation is untenable. Suppose that an instance of M causes an instance of M* by causing an instance of P*. By the physical realization thesis, which states that for a non-basic property, like a mental property M, to be instantiated in a system, "that system must instantiate an appropriate physical property, and further that whenever any system instantiates this physical property, the mental property must of necessity be instantiated as well" (Kim 1993a, 347) M itself will have a physical realization base P. On anyone's account the P is a cause of the P*, and there is a causal law that links the properties. What further role is there for the M-instance and the causal powers of M property, beyond those of the P and its P-property?

Kim argues that irreducible M instances of M properties don't do anything as particularly *M*-instances unreduced, for their causal powers do not go beyond those of their realization bases. This leads to the Causal Inheritance Principle:

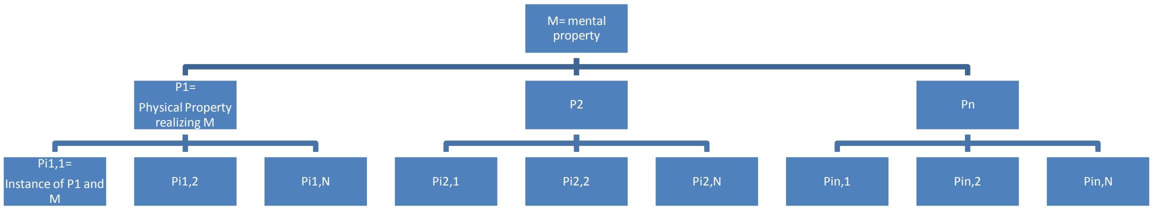
> **The Causal Inheritance Principle:** If M is realized on a given occasion by being realized by P, then the causal powers of this instance of M are identical with (perhaps a subset of) the causal powers of P (Kim 1993a, 355).

Suppose that the M had different causal properties than the P, Kim invites us to assume. The P would still cause P* and bring about M*, independent of M's causal properties. The exclusionary reasoning Kim employs tells us that the M, with respect to the P, is an epiphenomenon in the causation of the P*, the physical realization base of the M*; and that since the M would have to cause the P* in order to cause M*, M does not cause P. This is a general conclusion, for any putative mental causation case. Since the truth-condition of "M has causal powers" is that M-instances (the M's) have causal powers, mental properties do not have causal powers. Since this conclusion holds generally for all non-basic properties, of which mental properties are but an example, Kim holds, in consequence of his endorsement of Alexander's dictum that non-basic properties of things do not exist. In other words, Kim endorses *Epiphenomenal.*

In 1993 Kim is still committed to *Reduction.* According to Kim what the Causal Exclusion Problem shows is that we have a choice: we either reduce mental phenomena- properties, objects, and events- to physical phenomena or we adopt epiphenomenalism- which is independently implausible because of Alexander's dictum. Let us suppose, as Kim does, that the causal relata are events. Kim believes that in order for mental events to play a causal role, they must be identified with physical events. But Kim thinks that because of the phenomenon of multiple realization, according to which a mental property is realized in distinct sorts of physical realization bases, there are complexities in the reduction. He finishes the considered article by saying the type of reduction to be performed should take into account the multiple realizability of the mental. He says that information about such a reduction is found in his "Postscripts on Mental Causation," which was published that same year in the *Supervenience and Mind* anthology, and in his year-earlier "Multiple Realization and the Metaphysics of Reduction." In these essays Kim advocates identifying mental properties with the disjunction the physical properties that realize them. The way he develops this idea in terms of the Causal Inheritance Principle, is to say that:

> when we speak of the causal powers of M as such, we are speaking disjunctively of the causal powers of Pi's; and when we speak of the causal powers of a particular M-instance without knowing, or referring to, the specific Pi that realizes M on that occasion, we are again speaking disjunctively of the causal powers of M's many possible realizers (Kim 1993b, 363). Because of multiple realization, any given M-instance must be either a P1-instance or P2-instance or…, where P1, P2,… are realizers of M, and the set of all M-instances is the union of all these Pi-instances. In this sense, we may say that mental kind M is disjunctively identified with physical kinds P1, P2… Note that M is not identified with the disjunction of P1, P2,…; nor is an M-instance identified with an instance of the disjunctive property P1 or P2 or…. (Kim 1993b, 364).

We may call this proposal "multiple-type physicalism."



[49]

We should be clear about what the proposal is. It is not merely a token-token identity theory, even though the mental instances are to be identified with the physical instances that realize them. And Kim is not *just* saying that the mental property is identical with one of the members of the disjunction representing the physical extension of the mental kind. Rather than merely assert token-token identity though, it is a theory that acknowledges the fact that diverse *kinds* of physical properties realize M, not just non-identical physical tokens do (for example, at different times).[50] A way to understand the thesis, I think, is to think of the non-basic property of having shape. Having shape is a property realized by objects' having many kinds of shapes. A square has the property of having shape. And its property of having shape is realized by its having the property of being square. Kim wants to say that the square's having shape is,

[49] Diagram of Multiple Type Physicalism: instances have physical properties, which in turn realize the mental properties they have. Further, for any instance Pi1,1 having P1, which realizes M, Pi1,1's having M=P1.

[50] Multiple Type Physicalism aims to acknowledge the multiple realizability of mental properties and the idea that it is not sufficient for a property to be multiply realizable that it have more than one instance at different times. Multiple realizability requires that there be more than one compositional kind to have instances that realize the mental. It adds an extra layer of being as seen in Diagram 1.

in this particular case, identical with its being square: Its square shape, just is its shape and its shape is square. That is to say that the non-basic property of having shape, is reduced to its realizing basic property, being square in the instance in which an object is square. That is to say, Kim endorses *Reduction,* as well as *Real* and *Epiphenomenal* when he changes his view on whether epiphenomenal causal properties and relations are really causative.

Now: Kim endorses *Real, Epiphenomenal, Reduction*

Kim holds this view today as well. To support the thesis that Kim holds *Real, Epiphenomenal* and *Reduction* today, I shall be drawing material from *Mind in a Physical World*, the intermediate "Causal Drainage and Other Maintenance Chores With Mental Causation," the *Philosophy and Phenomenological Research* 2002-3 symposium on Kim's work on mental causation, and his most recent book *Physicalism or Something Near Enough*. I think that Kim's views have been sharpened in argumentative force and form of articulation, but have not only been fundamentally constant from at least 1993 but *a fortiori* are constant between these more recent pieces of work. He prefaces his latest book as containing "no startling new views about the mind-body problem beyond what can be found in (his) earlier book *Mind in a Physical World*," but adds only "better focused and motivated arguments" (Kim 2005, xi). From this bit of evidence and the observation of real constancy in the views expressed, I will use material freely from all three sources to support the thesis that Kim holds *Real, Epiphenomenal* and *Reduction* now, i.e., that these works represent his current views.

In *Mind in a Physical World* he says that "*macroproperties can, and in general do, have their own causal powers, powers that go beyond the causal powers of their* *micro-constituents.* This is an obvious but important point to keep in mind." He does in fact say that "a neural assembly consisting of many thousands of neurons will have properties whose causal powers go beyond the causal powers of the properties of its constituent neurons, or subassemblies…" (Kim 1998, 85). These comments are explicit enough to prove that Kim can be attributed the belief in *Real* now.

In all the cited works Kim employs the Problem of Causal Exclusion. This problem is designed not to show that mental properties and their instances are epiphenomenal with respect to physical properties and their instances. Rather, it is to show that mental properties and instances must either be reduced to physical properties and instances, or they are epiphenomenal. Kim wants to force the choice between "reduction or causal impotence" (Kim 2003, 165-166). This implies that mental properties and their instances that are not reduced to physical properties and their instances would have causal powers that "go beyond the causal powers of their microconstituents," which is decidedly implausible, Kim thinks, and therefore must be causally impotent on their own.

There is however, a step we further need to justify this claim with and that is to say that it is because mental properties and instances are non-basic and non-basic things are not irreducibly causally potent apart from their bases, that Kim makes this judgment. This gets us into the debate about whether the Problem of Causal Exclusion generalizes to all other non-basic properties. To make the claim that Kim is committed to *Epiphenomenal* now to work, I need the premise that what he is claiming about mental properties and instances is so because they are not basic; not because there is some peculiar feature of mentality in virtue of which all other non-basic features of reality get their causal autonomy, while the mental does not. I believe Kim is unsuccessful in his attempt to isolate the mental in this way.

Relevantly, there is an immediate question which Kim does not properly address, in my view. Why cannot mental properties be macro-properties with their own distinctive causal powers? In 1984 he thinks that “macrocausation should be viewed as a kind of epiphenomenal causation…; that macrocausation as epiphenomenal causation should be explained as “supervenient causation”…; and that psychological causation…is plausibly assimilated to macrocausation- that is, it is to be construed as supervenient epiphenomenal causation” (Kim 1984, 95). Up until this point he has made reference to macrocausation, the phenomenon that links events in virtue of non-basic properties, as a kind of “causation” that is epiphenomenal. Under this model mental causation was to be understood as on a par with macrocausation. In 2005 Kim admits that the Problem of Causal Exclusion, which depends on there being a causally closed domain, presumed to be the bottom microphysical domain applies to biological causation, chemical causation, geological causation, “and the rest” (Kim 2005, 66).

And he assumes that such a basic domain is the domain in which the “world looks like this: all the things that exist are physical things- either basic bits of matter or wholly made up of bits of matter. These physical things have properties. What properties? First, there are basic physical properties, like mass, size, shape, electric charge, and so on-properties and magnitudes in terms of which laws of physics are formulated” (Kim 2002, 640). And he alludes again to Edwards in his 2003 article, formulating a principle in his echo:

> **Edwards' dictum:** There is a tension between vertical determination and horizontal causation. In fact, vertical determination excludes horizontal causation.

But in the quote above he says that macroproperties have their own causal powers (Kim 2003, 153).[51] And in the "Précis to his Mind in a Physical World" he says that "thoughts and pains are determined by biology, and ultimately physicochemistry…" (Kim 2003, 641) ignoring his commitment to the basic level as the only level in which causal powers reside, for physicochemistry is not ultimate, but wholly determined by, and epiphenomenal if unreduced to, the properties of particles of the Standard Model, if Kim's exclusionary reasoning is correct. The exclusionary reasoning is grounded on the assumption expressed in Edwards' dictum. Given that causation does occur, it is confined to the inhabitants of the fundamental level of physics upon which all else supervenes and "rides for free." In support of his basic fundamentalism he begins his reply to Frank Jackson by saying "that the only consistent and robust form of physicalism, something worth aiming at or arguing about, is reductionist physicalism" (Kim 2003, 671).

We may conclude that Kim thinks that either *both* that irreducible non-basic macroproperties and their instances are epiphenomenal with respect to their corresponding realizing micro-properties and their instances *and* that irreducible non-basic macroproperties are not epiphenomenal with respect to their realizing properties and their instances (a contradiction, in effect),[52] on the one hand, or that non-basic

[51] The dictum is coupled with the comment that when he first quoted Edwards in his 1984 article he did not realize the full significance of this consideration for the Causal Exclusion Argument.

[52] Since he believes in causal "physico-chemical properties"- as well as thoughts and pains- as said in his last paragraph, and he also believes that only things at the fundamental level are causal.

properties and their instances are epiphenomenal with respect to their realizing properties and instances, on the other.[53] Therefore, Kim believes that non-basic macroproperties of things are epiphenomenal with respect to their realizing basic properties of things. However, since he also holds Alexander's dictum to be true, he must say such properties do not exist. Therefore, Kim is committed to *Epiphenomenal.*

Kim is also committed to the idea that non-basic properties of things are identical with basic properties of things, *Reduction*. The Problem of Causal Exclusion shows, for Kim, that there must be token-token identifications between physical and mental causal relata, and that the properties exemplified in the relata must be identified[54] in order to secure mental causation. In support of the token-token identity claim he thinks the "If M is to retain its causal status, it must be reducible to P-at least, the given instantiation of M must be reductively identifiable with the instantiation, on that occasion, of its supervenience or realization base" (Kim 2002, 642). But are mental properties physically reducible? He answers that he has "argued that if they are to be causally efficacious, whether with respect to physical properties or other mental properties, they must be reducible to physical properties" (Kim 2002, 643). In his most recent book, he argues for the same thesis and rearticulates his model of reduction. According to his model, there are three steps in making reductions. First, we take the conceptual step of defining the non-basic property as a functional property "in terms of the causal work it is supposed to perform" while knowing that "the phenomenon

[53] Since he believes *non*-basic properties exist, but are pre-empted from entering into causal relations of their own.

[54] By disjunction elimination of the already reduced to absurdity disjunct; also if the contradictory disjunct is somehow accepted, by conjunction elimination, the conclusion follows; also even if the second disjunct were eliminated, the conclusion follows from the contradictory remaining disjunct, since anything follows from a contradiction.

involved to be reducible to its physical realizer." The second step is to employ the scientific method to tell us what the realizer of that functional property of the object in question is- that is, the basic properties with which the non-basic functional properties are identical. And the third step is to develop an explanation of the mechanisms by which the causal work is done by the basic physical realizers (Kim 2005, 164). That is to say, Kim accepts *Reduction* now.

In this section I have gone through the significant stages in Kim's thinking about causation of higher-level non-basic things, trying to identify his views on theses *Real, Epiphenomenal* and *Reduction* now. My conclusion is that all three theses are constantly attributable to him from 1984 until the present.

2.3.3 The Tension Between *Real, Epiphenomenal,* and *Reduction*

The tension should be clear. Either irreducibly non-basic properties of things are causal or they are not. If they are not, then by Alexander's dictum, they are not real and so mental properties are not real and *Real* is contradicted. If they are, then non-basic properties, including mind, are not threatened by elimination. But how do they deal with the Problem of Causal Exclusion? Kim argues that they must be basic properties in the last analysis. But can non-basic properties be basic properties? The answer that they can't seems to be evident, just by the law of non-contradiction. But then, if they are distinct, it follows from *Epiphenomenal* that the non-basic ones are epiphenomenal. If so, then non-basic properties do not exist, by Alexander's dictum. But epiphenomenal properties cannot be identical with causal properties, as would be required by *Reduction. A fortiori* (because they would be eliminated by Alexander's dictum) non-existent properties cannot *be* existent properties.

In commenting on Kim's 1998 book, however, Jackson notes that Kim's way of making the identification is to opt for a sparse conception of properties, under which predication is not a property-forming operation, and to interpret mental predicates as mere designators of real properties, the basic ones, members of the mental categories in virtue of playing a specified role (Jackson 2002, 646). In that case we say "M=P" is true when P plays the role definitionally given by "M." If we generalize this strategy for the rest of higher-level non-basic properties that might countenanced, such as macroproperties, *Real* is contradicted.

It should be clear that a solution to the Problem of Causal Exclusion must not simultaneously endorse *Real, Epiphenomenal,* and *Reduction.*

2.4 Non-Reductivist Solutions to the Problem of Causal Exclusion

In their recent article on mental causation, “The Metaphysics of Mental Causation,” Cynthia and Graham Macdonald propose an alternative to Kim’s solution to the Problem of Causal Exclusion.[55] They use the framework of Cynthia Macdonald’s version of the Property Exemplification Account (PEA) of events, developed in her *Mind-Body Identity Theories,* to do so. As Graham Macdonald pointed out in the conference on Emergence at Queen’s University Belfast in April 2007, since the account was developed by 1986, when the Problem of Causal Exclusion had yet to be taken into account and does provide a solution to the problem, the Macdonalds’ theory is imbued with a confirmational boost characteristic of successful novel theoretic prediction. That is to say, the theory has a proper application to an area for which it was not designed, in that it was constructed *ab initio* and not *ex post facto,* and therefore warrants us in being more confident of its truth. However, I think there are shortcomings with the account, and this is not the only account that provides a solution. The other proposed solutions involve the trope theory, the constitution theory, and Kim’s reductive materialism.[56] I have already discussed Kim’s reductive materialist solution and will consequently only discuss the non-reductivist alternatives. After explaining the problem, as the Macdonalds set it up, and explaining their solution, I will argue that the trope theory is not in competition with the PEA of events relevant to that solution. Everything the Macdonalds want the PEA to do, can also be done with trope

[55] In the following sections I will formulate the problem and solutions using the physical/mental dichotomy that is traditionally used to do so.

[56] Jackson and Pettit (1990) accept that mental properties are explanatory but not causally determinative of effects. Though see Cynthia Macdonald and Graham Macdonald (2007).

theory to underlie it (rather than the universalist conception of properties). Therefore, the Problem of Causal Exclusion gives us no forced choice between trope theory and the Macdonads' solution. I will then articulate the constitutional account developed by Derk Pereboom, modifying it to account for the plausible assumption that it is events, rather than objects, that are the relata of the causal relation. This is, I think, the relevant contender.

2.4.1 Generating the Problem of Causal Exclusion Again

Recall (2.2.7) *The Mental Comes Back to Power.* There we went through each premise in Kim's Problem of Causal Exclusion and saw that our notion of the physical, independently argued for, dissolves much of the apparent problem. However, the problem can be resurrected if we do two things. One, we assume that mental properties of objects are non-basic and we substitute "physical" for "basic," in order to understand the supervenience relation as a relation of dependency of non-basic properties on basic properties. (For expositional ease, I will continue to speak of the physical to follow the verbal tradition.) And secondly, if premise 3 of Kim's argument is maintained to be adequately replaced by 3', so as to be left with the problem of saying whether it is option a) or b) that explains the instantiation of a mental property M2. We had revised the premise in Kim's argument in order to give it more force. The premise now looks like this:

3' M2 is instantiated on this occasion: a) because an instance of M1 caused an instance of M2 (as an instance of supervenient causation), or b) because an instance of P1, the M1's instance's physical supervenience base, causes M2 to be instantiated (perhaps by causing P2, M2's physical supervenience base, to be instantiated).

Put like this, the question that arises is: How is it that an instance of M1 causes an instance of M2 in virtue of ("because" of) being an M1? That is, the *qua* problem remains; the problem of saying how it is that the mental can cause *qua* mental. A plausible answer must also deal with the assumption that mental properties, while physical in that they are exhibited in space-time and are regulated by the laws of nature, are dependent on more basic properties of objects as implied by their being supervenient on other properties.

There are four theses minimal physicalists might be committed to, while also endorsing mental causation. Minimal physicalism is the view that although each mental event or phenomenon is a physical event or phenomenon, mental properties are wholly distinct from physical ones (Macdonald, C. and Macdonald, G. 2006, 541).

The endorsement of mental causation is embodied in the following claim (Macdonald, C. and Macdonald, G. 2006, Section 1).

Mental Causal Relevance (MCR) Mental properties of physical events are causally relevant to the physical effects those events bring about.

The endorsement of physicalism is embodied in the following three theses:

Physical Causal Relevance (PCR) Physical properties of physical events are causally relevant to the physical effects those events bring about.

Exclusion (EXCL) If a property, *P*, of a cause, *c*, is causally sufficient for an effect, *e*, then no other property, *Q*, distinct from and independent of *P*, is causally relevant for *e*.

Closure (CLOS) If a physical event or phenomenon has any cause, it has a sufficient physical cause, whose physical properties are causally sufficient for its effect.

MCR expresses the commitment to mental causation as the basis for mental explanation. Mental Realists want to retain several assumptions that presuppose MCR. We want to say that the folk psychological explanation of behavior, which cites reasons, beliefs, purposes, perceptions, desires and other mental phenomena, makes reference to things that cause the behaviour to be explained, and that the kind of phenomena referred to is causally relevant. We also want to say that the field of psychology proper is a scientific discipline which purports to theoretically reveal those scientifically interesting properties of mentality that explain the behaviour of mental beings. And we want to say that the mental causes *qua* mental, i.e., that it is in virtue of its mental properties that (at least many) mental events cause. These are not uncontroversial or universally endorsed desiderata, but it is clear that it is only if MCR is true that these popular desiderata can be satisfied. Further, because mental causation presupposes MCR, we want a solution to the Problem of Causal Exclusion.

The Macdonalds understand PCR to mean that “physical properties of physical events are such that their instances are causally effective in bringing about physical effects of those events” (Macdonald, C. and Macdonald, G. 2006, 544). It is clear that for physical properties to be causally relevant they must be causally effective, since any property that is causally sufficient for some effect must be effective in bringing it about. However, as the Macdonalds are careful to point out, not all causally efficacious properties are causally relevant, and therefore PCR is stronger. It says that instances of physical properties are not only causally effective, in being properties of causally effective events, but are also relevant- or "salient" (Macdonald, C. and Macdonald, G. 2006, 568). Relevance is a pragmatic phenomenon, as Grice and van Fraassen have recognized in similar contexts.[57, 58] In some contexts, physical properties of physical events are irrelevant in the causal explanation of why a physical event occurs. For example, take the scenario in which the four fundamental forces acting upon a particular collection of quarks and electrons I picked up from the grocery stores from which my food has come throughout my life are causally sufficient to get me to write this very sentence. In most contexts where we would ask for a causal explanation of this event (my writing), if the explanation is couched on the properties of quarks, electrons and fundamental forces, the explanation would not be a relevant one.[59]

From our analysis of Physicalism and what it is to be a physical property (Section 2.2.6), we see that mental properties, given that they are exhibited in space-time and are governed by the laws of nature, must be physical properties. We have already seen how, with this conception at hand, the Problem of Causal Exclusion

[57] H.P. Grice's theory of Conversational Implicature of (1975) has as a contextual rule followed in conversation called the Relevance Maxim. The rule says that, in conversation, people ought to say relevant things, and can be expected to do so. Van Fraassen (1980) has forceful illustrations of the role of relevance in the pragmatics of explanation. Read on for an example. Further, one should not think that because causal relevance has a pragmatic factor, that properties that are causally relevant or explanatorily relevant have their causal powers in virtue and solely in virtue of our interests. Properties have these things even without our knowledge of them. Having a pragmatic factor figuring in counting a property as being causally relevant does not imply being irrealist about that property.

[58] Some may want to object that while relevance is context-sensitive, this is a different matter from its being a pragmatic matter. They may say that while explanation is interest relative, and causal relevance is interest relative, causal relevance does not collapse into a variety of explanatory relevance, because more is involved in explanation than just causal relevance. For one thing, an event might have more than one causally relevant property. However, pragmatic factors just are the context-sensitive factors of explanation. If there is more to the explanatory relevance of properties cited in a causal explanation (pertaining to a variety of explanatory relevance) than the properties that are causally relevant, then it cannot be *because* events may have more than one explanatorily relevant property, for likewise, events can have more than one causally relevant property. Thanks to Cynthia Macdonald for articulating this objection.

[59] This is important for the use of causal models *a la* Hitchcock (2001).

largely melts away. If Physicalism is true, then having a mental property is a way of having a physical property, in the way in which having an apple is a way of having a fruit. For those who do not accept my analysis of Physicalism and physical properties, the implicit separateness of the physical and the mental in much of this debate will allow them to formulate the problem with more ease. However, the part that has not melted away for me, as noted, was a way to deal with the issue of how mental properties can be causally relevant to an effect, when they depend on their supervenience base. However, in order to even talk with non-adherents to my version of Physicalism, we need to understand what they mean by “physical” in a way that does not make the problem a non-starter. The way to do that is just to suppose that “physical properties” refers to the supervenience base of mental properties.

EXCL comes from Kim’s Principle of Explanatory Exclusion, which states, “No event can be given more than one *complete* and *independent* explanation” (Macdonald, C. and Macdonald, G. 2006). As the Macdonalds point out, Kim leaves the notions of explanatory completeness and independence undefined. However, it is clear that complete and independent explanations “cannot be identical; they cannot” just cite “partial causes, constituents of a larger cause; they cannot be different links in the causal chain leading to e_1; neither cause can be a ‘part’ of the other cause” (Macdonald, C. and Macdonald, G. 2006, 544). The principle also descends from Yablo’s exclusion principle stating that “If a property X is causally sufficient for an event *y*, then no property X^* distinct from X is causally relevant to *y*.”[60] Yablo objects that if properties were “efficacious not absolutely, but only relative to some specified

[60] This is derived from Yablo (1992, 247), in line with his instructions (in his fn. 5) for translating the event-version into the above property-version of *Exclusion*.

effect" this would be unsatisfactory (Macdonald, C. and Macdonald, G. 2006, fn. 14). As the Macdonalds point out Yablo conflates efficacy and relevance. For, properties of the cause-event c sufficient for y are all absolutely sufficient for y, given that c is an instance of all these properties (which are all causally sufficient given that c is causally sufficient for y),[61] but it is only for some contexts or kinds of contexts that any such causally sufficient property is causally relevant. This would indeed make causally sufficient properties relevant only relatively, but this is a basic fact that has to be accepted given that we have selected *relevant* properties, given that pragmatics, at the core of relevance, is essentially context-relative. A further criticism of Yablo's objection is in order. For while it is true that the specification of the effect of interest does relativize which properties are causally relevant, it is not only relative to the effect, but also the interests in the field of discourse that make the effect in question relative to it. Van Fraassen illustrates some of the pragmatics of explanation in an example I will adapt to the present debate (van Fraassen 1980, 127). Suppose we want to explain why Adam ate the apple. If EXC is at all plausible, there is an event with a property P such that P is causally sufficient for Adam's eating the apple (omitting time for simplicity). But in different contexts different properties of the effect, and those that cause it, become relevant. For example if the interest is in why *Adam*, rather than the trickster snake, ate the apple, properties of the sufficient cause, like it being such that the snake wanted to trick Adam, become relevant. But if the context is in explaining why Adam ate the *apple*, rather than any pear, other properties of c become relevant.

[61] The reader may notice the background assumption that property instances are the things that have them. This is a part of the Macdonalds' non-reductive monism; explicitly articulated in Macdonald, C. and Macdonald, G. (2006, fn. 30).

Another point that is worth mentioning is that EXCL can do the job it is supposed to, namely generate the Problem of Causal Exclusion, only if its scope is not unrestricted. For surely there are overdetermined events which have different independent and causally sufficient properties of their distinct overdetermining causes which are independently causally relevant. To adapt an example from van Fraassen again, suppose there is a car crash which is overdetermined by the road's being soapy (plus other background conditions x) at t and the STOP sign's being covered (plus x) at t. In that case, the road's being soapy (plus x) at t is sufficient for the car crash. And independently of this sufficient cause, the STOP sign's being covered (plus x) at t is also a sufficient cause of the crash. In the case we are considering, because the causes overdetermine the car crash, the car crash would have been brought into existence had only one of the considered causes (plus x) been present. Surely, the properties of being a soapy road (plus x) and being a covered STOP sign (plus x) are each causally relevant and independent and distinct from one another. There are therefore (at least) two independent causally relevant properties sufficient for the effect in question. Or to use a favorite of Kim's, the particular distinct direction-properties of two death-overdetermining bullets are surely distinct and independent of one another and yet causally relevant toward the effect.

Perhaps a way making this case not falsify EXCL from the beginning is to suppose that a causally sufficient P *must* be the only causally relevant property, where we stipulate from the beginning that we are not considering these kinds of overdetermination cases as part of the scope of EXCL.

A potential problem with this move is that it trivializes EXCL and robs it of its ability to generate the problem. Restricting the scope of EXCL in this way has the effect of letting us say that if mental causation implies overdetermination, EXCL does not apply to it- so systematic overdetermination is allowed in mental causation, which strikes many as magical as there always being an overdetermining killer for every kill actually performed by another killer. Or if EXCL applies to the mental causation cases, then the threat of such implausible systematic overdeterminations are excluded *ab initio.*

Another potential way of trying to make EXCL independently plausible is to amend it to:

> **EXC!**: If a property, *P*, of a cause, *c*, is causally sufficient for an effect, *e*, then no other property, *Q* of c, distinct from and independent of *P*, is causally relevant for *e*.

Here we have made P and Q be attributed to the same cause c so that the case is not refuted by the case of independent bullets from independent killers overdetermining the killing of a man. However, EXC! is still deficient on quite similar grounds and we can see this by just tweaking a bit with the scenario. Suppose the killer fires two bullets sufficient for killing the victim. Name those bullets "Bob" and "Sally." Then there are two properties of *the shooting* by the killer, namely, being a shooting of Bob and a shooting of Sally, which are distinct and independent of one another. And yet both are properties of the cause, sufficient for the killing. Thus, EXC! is still independently implausible.

EXC has another fundamental problem. It is that the Problem of Causal Exclusion is a problem that is generated, not because there are distinct and independent mental properties of the causing event of which we wish to say that it is in virtue of these mental properties that the event causes, but rather, it is the following. It is because mental properties of the event are supposed to be dependent, by being supervenient, on some other properties of the event, and it is this putative dependency that creates the mental epiphenomenalist scenario that Kim describes like this:

7. The P1-instance causes the P2-instance and the M1-instance supervenes on the P1-instance and the M2-instance supervenes on the P2-instance.

It is precisely because what is mental in the envisaged scenario is seen to depend, at any time, on some other fact, and because the sufficient causal connection is perceived to be at the level on which the mental depends, that a threat to MCR is generated.

Perhaps EXC should read (ignoring previously mentioned problems):

Exclusion* (EXC*) If a property, *P*, of a cause, *c*, is causally sufficient for an effect, *e*, then no other property, *Q*, distinct from and *dependent* on *P*, is causally relevant for *e*.

CLOS is a central thesis of Physicalism. There is no caused physical event whose causal ancestry is not physical, in virtue of which the event in question comes about.

This does not, of course, imply that, for whatever "physical" might mean here, physical events and their properties cannot, besides being physical, be other things, like biological or mental. It is left open that indeed biological and mental events are physical. Otherwise, Physicalism or Mental Realism would be eliminated prior to any further investigation. Further, CLOS does not assert determinism. Rather, it says that the causes of caused events are physical causes.

The Macdonalds connect the theses to generate the problem thus:

> How does CLOS work, together with PCR, and EXCL just mentioned, to generate the charge that MCR is false? Well, we know from MCR, PCR, and EXCL that if we have two properties (one mental and one physical) of a cause, each of which can be cited in two causal explanations (one mental and one physical) of a single explanandum, we cannot (by EXCL) accept both unless neither of those properties, taken alone, is causally sufficient. What CLOS entails is that there is some causally sufficient (physical) property (Macdonald, C. and Macdonald, G. 2006, 546).

It appears as though the mental property is preempted from making a causal contribution and thus be able to figure in a causal explanation of the physical event effect, and MCR is ruled out.

2.4.2 Non-Reductive Monism

According to the Macdonalds, non-reductive physicalism about the mental is saved from the Causal Exclusion Argument if they can maintain that mental properties have causally efficacious instances whose mental properties are causally relevant, Non-Reductive Monism. To develop this idea the Macdonalds develop a version of the Property Exemplification Account (PEA) of events originated by Jaegwon Kim. According to this account events are structured particulars, whose components are constitutive objects, properties, and times. These components are not related mereologically. Rather, one component exemplifies another at yet another. That is, any event is composed of an object which exemplifies a property at a time. We can formulate the core of this view of events in two theses:

Existence Condition: Event [*x*,*P*,*t*] exists if and only if the object *x* has the property *P* at time *t*.

Identity Condition: Event [*x*,*P*,*t*] is identical with event [*y,Q,t'*] if and only if the object *x* is identical with the object *y*, the property *P* is identical with the property *Q*, and the time *t* is identical with the time *t'*.

P and Q are variables ranging over properties that are constitutive of events. Constitutive properties of events are those properties of events that events have essentially, without which the events in question would not be the events they are. Apart from constitutive properties, events also have characterizing properties. Characterizing properties of events can be determined by the properties of the object, but are not the same properties. For example, suppose as Kim (but not the Macdonalds) would have it that the property, has pain, is a constitutive property of the object of the event which is identical with that event exemplifying being a having of a pain now (Macdonald, C. and Macdonald, G. 2006, 556-557). The Macdonalds theorize that "to say that mental events are identical with physical events is to say that each event which is (= is identical with) an exemplifying of a mental property of a subject in that subject at a time is identical with an exemplifying of a physical property of that subject in that subject at that time. Crucially, this amounts to the claim that there is just *one exemplifying* of two properties, one mental, and one physical, by an object at a time" (Macdonald, C. and Macdonald, G. 2006, 560-561). This claim is made plausible by the idea that such a co-instantiation of properties happens in the determinate-determinable case. Being red is a determinate of being coloured, and a thing that exemplifies being red at a time exemplifies being coloured, and yet the two properties do not compete.

Similarly, an event that is an exemplifying of being red is an exemplifying of being coloured.[62] The idea of co-instantiation of properties in a single instance is formalized thus:

Co-Instantiation (CI) Two or more properties of an event can be co-instantiated in a single instance, i.e., there can be just one instance of distinct properties (Macdonald, C. and Macdonald, G. 2006, 562).

The properties in virtue of which events cause are not the constitutive properties of the objects whose exemplifying of them at times, are those events. Rather, events cause in virtue of *their* properties, not their objects'. Further, higher level properties (e.g. being a change in color) of events are had just by the event's having a lower-level property (e.g. being a change from yellow to green). We can formulate a stronger thesis than indicated in CI to pick out the sort of property dependence at issue.

Property Dependence PD When properties of events are related as higher-level to lower-level and the lower-level properties realize the higher-level ones, an event exemplifies the higher-level one just by exemplifying the lower-level one (Macdonald, C. and Macdonald, G. 2006, 564).

The Property Dependence thesis tells us that, on the assumption that mental properties of mental events are higher-level properties of physical-lower level realizing properties of the same event, the event has the mental property it does just by having the lower-level physical property it has. For example, an event has the property of being a thinking of Vienna just by having the property of being a neural event β. For the Macdonalds, however, the sort of realization relation at issue between higher level

[62] The intuitive idea that events are fundamentally changes requires that events be exemplifications of more than one property, since change involves the having of a static property and then having another. The Macdonalds endorse this view, inherited from Lombard (1986) .

mental properties and lower level physical properties of events is akin to, but distinct from, the determinate/determinable realization relation. The latter involves a conceptual necessity, while the former does not necessarily involve one. It may be a homological necessity that arises out of the metaphysically contingently actual laws of physics. For them, the relation is given in Strong Supervenience.

Strong Supervenience *(SS)* *M*-properties strongly supervene on *P*-properties =*do*.

> For any possible worlds *w* and *w**, and any individuals *x* and *y*, if *x* in *w* is a *P*-twin of *y* in *w**, and the actual world's laws of physics hold in both, then *x* in *w* is an *M*-twin of *y* in *w** (Macdonald, C. and Macdonald, G. 2006, 565).

In any strong supervenience case, there is just one event (the relevant individual), one instance, of two properties. Just like determinate/determinable properties of events don't compete for causal sufficiency, neither do lower level subvenient properties of events compete with higher level strongly supervenient properties of the event. This secures the idea that mental properties of physical events are causally efficacious. Therefore, the possibility that mental properties of physical events are causally relevant is not ruled out.

However, causal sufficiency is a necessary but not sufficient condition on causal relevance. On this score, as they point out, the causal relevance of the mental is made plausible by noticing several things. First, that there is no special problem for mental causal relevance that is not present in all other higher-level causal relevance cases. The mental is just one example of the ubiquitous phenomenon of higher level causal relevance. And secondly, higher-level causal relevance, including mental properties, fits in well with the various ways of filling in the schema for causal relevance in the literature on causation:

> **CF** C has an effect on *E* if *E* depends upon *C* – that is, if *E* varies as *C* is varied while holding fixed other appropriate factors (Macdonald, C. and Macdonald, G. 2006, 572).

Two prominent ways of filling out the schema CF, guided by the Principle of the Homological Character of Causation, for example, are Lewis' counterfactual dependence and his later counterfactual co-variation accounts, and Hitchcock's (2001) "two-cause" causal models. The Principle of the Homological Character of Causation states that causal relations instantiate natural laws "covering" the events. The Macdonalds note that there will be very many physical properties of any one event which do not figure in such homological connections, so the problem of causal relevance is quite general, applying not only to mental properties, but also to chemical, biological and even some physical properties of the events in question themselves (Macdonald, C. and Macdonald, G. 2006, 568). Once one understands, however, that properties are said to be causally sufficient when their instances are, we understand that causal sufficiency comes very cheap but is not the only factor that determines relevance. Causally relevant properties require efficacy, but also the being able to figure in appropriate regular patterns, whose fundamental nature is indicated as being something like what theorists on causation more generally tell us. According to the understanding available in this literature, mental causal relevance has no special problem at all, and causally relevant mental properties fit into this understanding smoothly.

Non-Reductive Monism is compatible with EXCL, asserting that while there may well be more than one causally relevant property Q distinct from the causally sufficient property P, of cause c, Q is nevertheless dependent on P. Such a dependency would be the way that the Property-Dependence thesis indicates. But given that P is a causally sufficient property of c for e, and Q must be as well (given that c is identical with this P-instance and is identical with this Q-instance, and that to say "P is causally sufficient" is interpreted as "P-instance is causally sufficient"), there is no reason to believe that because any properties of c are causally sufficient, that some other properties of c are not. Given that all the properties of a thing that has them are instanced in the single instance that is that thing (in this case an event), all its properties are causally sufficient for whatever effects that thing has. Some of these causally sufficient properties of the thing in question will be causally relevant and some won't, and no property can be causally relevant unless it is causally sufficient. However, there is no entailment from c's having causally sufficient property P to which properties of c will or will not be causally relevant.

In terms of the way Kim set up the Problem, recall proposition 3: M2 is instantiated on this occasion: a) because an instance of M1 caused an instance of M2, or b) because an instance of P2, the M2-instance's physical supervenience base, is instantiated. As we saw, there is no competition here, since a) is considered as a causal explanation of M2, and b) is considered as a reductive explanation of M2. This is seen by the fact that the instance of M1, the considered cause of the instance of M2, is taken to happen *before* the instance of P2, the supposedly competing factor in the generation of the instance of M2. But factors that occur at different times may well just be factors in a causal chain leading to the considered effect, M2, in this case. Thus, let us re-state more compellingly what the two completing explanations are supposed to be. To be fixed proposition 3 in the Causal Exclusion Argument becomes:

Premise 3*: M2 is instantiated on this occasion: a) because an instance of M1 causes an instance of M2 or b*) because an instance of P1, M1 instance's physical realization base, causes an instance of M2 (perhaps by causing the instance of P2, the M2's physical realization base).

The "because" in a) and b) is ambiguous between whether the problem is that of specifying a sufficient cause, or whether the sufficient cause is to be referenced in terms of its causally relevant properties as well. According to the Macdonalds, the instance of M1 and P1 are the same. There is one instance of two properties, and that instance causes the instance of M2, which is itself identical with an instance of P2. In this way, there is no competition between M1 and P1. Now if what is demanded is that the mental properties in question be causally relevant as well, then we need to look at whether varying the mental properties of the M1-instance would vary the properties of the M2-instance, holding other appropriate factors fixed (as CF requires). Here, the Macdonalds argue, the overwhelming number of models of causation (which they think particularly model causally relevant dependence relations) support the thesis that mental properties are causally relevant.

Let us consider the version of the Problem of Causal Exclusion that we saw in section 2.1.2 where it was applied to anomalous monism. Here we were lead to the conclusion that:

6*. The P1-instance causes the P2-instance and the anomalous M1-instance supervenes on the P1-instance and the M2-instance supervenes on the P2-instance.

This proposition creates the impression that the connection between the M1-instance and the M2-instance is epiphenomenal with respect to the connection between the P1-instance and the M2-instance. The Madonalds argue that the M1-instance and the P1-instance are identical, and similarly for the M2-instance and the P2-instance, so there is no competition between instances, so there is no question of whether the M1 instance is causally sufficient for the M2 instance. However, the question arises as to whether, given that causality is nomologically governed, there are laws that connect the properties M1 and M2, and therefore whether the causing instance causes *qua* mental. The Macdonalds hold that the instance does cause *qua* mental because the mental property is causally relevant.

Is being causally relevant sufficient for being lawful? That is, are the properties of being causally relevant and the properties that figure in causal laws the same properties? If so, then the solution is apparently effective. In conceptual tandem with this supposition, Cynthia Macdonald considers changing the Principle of the Nomological Character of Causation to the Principle of the Nomological Character of Causally Relevant Properties, for “these properties, the ‘causally relevant’ ones, are the prime candidates for properties that might be in laws linking such events” (Macdonald 1989, 159). For if not, it looks like the Problem of Causal Exclusion re-emerges.

This solution is also not compatible with EXC*, which the Macdonalds reject.

2.4.3 Does Trope Theory Compete?

The Macdonalds defend the idea that the universalist conception of properties is an essential part of the PEA of events, which the Macdonalds use to solve the Problem of Causal Exclusion (Macdonald, C. and Macdonald, G. 2006, 548). The Platonist universalist account of properties they endorse holds that properties are abstract entities existing outside of space-time and are entities that do not in themselves partake in the causal structure of the world. According to this account, properties are multiply exemplifiable entities. When something has a property, the property exists independently of the thing that has it, and any two things that have it exemplify the same entity. When two things are red, they both exemplify the same entity: redness. Further, when we have higher-order predication, we speak of the abstract entity in the subject-phrase. Thus, when we say that red is a color, we say something about redness itself, rather than the things that are red.

However, as can be seen, I have formulated their solution independently of talk of universals or tropes (ways of being). This is a *prima facie* case for the idea that the PEA of events does not entail a universalist conception of properties, and therefore that the universalist conception is a separate thesis upon which the Macdonalds' solution to the Problem of Causal Exclusion does not depend. The PEA itself can be given a trope-theoretic interpretation. Therefore, it is worth seeing whether the trope theory survives their criticisms that it does not adequately solve the Problem of Causal Exclusion.

According to the classic trope theory, tropes are the foundations of all things.[63] Tropes are particular things. Any red thing has a quite particular redness, and saying that two things are red is like saying that two soldiers have the same uniform. Also, for

[63] Originating in Williams (1953) and also expressed in Heil and Robb (2003).

higher-order predication, saying that the property red has the property of being a color, is saying that all particular things that are red, have color.[64] Tropes are bound by resemblance relations to other tropes, which together, form trope property classes. In the trope conception, property types are classes of tropes. Under the trope conception, events are tropes or complexes of tropes. If we like the PEA of events, then the constitutive objects, properties, and times, can themselves be tropes.

It may at this point be doubted that trope theorists can avail themselves of the notion of instantiation, or exemplification, as the universalist can. If they cannot, then a trope-theoretic interpretation of the PEA of events is not forthcoming and a solution based on the PEA of events may compete with trope theory. An example of the sort of ideas evoked to argue for the universalist conception of properties is found in Plato's *Parmenides*: "there exist certain Forms of which these other things come to partake and so to be called after their names; by coming to partake in Likeness or Largeness or Beauty or Justice, they become like or large or beautiful or just."[65] Loux discerns that what is proposed is a general schema for attribute agreement, or for our purposes, propertyhood: "where a number of objects, a…n, agree in attribute (property), there is a thing ß, and a relation, R, such that each of the a…n agree in attribute (property) by being all beautiful, or just, or whatever" (Loux 2002, 22; parentheses added). The relation R is Plato's "partake" and today's "exemplify" or "instantiate" since the relation R connects the thing that has the property that other things may also have, with that property that other things may also have, like being beautiful or just or whatever.

[64] Relevantly, saying that a trope property type has causal powers is saying that each member of the corresponding class has causal powers.
[65] *Parmenides* 130E-131A quoted in Michael J. Loux (2002, 21). There are uncertainties as to whether Plato himself held the "Platonist view," which I will ignore.

This is all fine for the trope-theorist. For s/he can indeed endorse the schema Loux proposes. S/he can say that when more than one thing a…n has the same property, there is a thing ß, that is a set that contains a…n; and there is a relation R that is the class-membership condition characteristic of ß, such that each a…n meets it (this is how each member resembles the rest). Meeting that condition by having a given trope is said to be "instantiating the property" or "exemplifying the property." Each a…n is an instance or example of the having of the property they each have in virtue of which they are each a member of the common set ß. For example, the trope property class of being red has all the red things as members. Each member instantiates being red, and it is in virtue of the fact that each member satisfies the set membership condition characteristic of the redness class that it is a member of the set of all red things. This is not to say that there is no difference between universalism and trope theory, but just that the two have the instantiation and exemplification available to them, though they differ in their account of what instantiation and exemplification consist in.[66]

If it is objected that properties are multiply exemplifiable but that trope classes are not, we can argue that they are (in the required way) by pointing to the following. First, it should not be required that trope property classes always be multiply exemplifiable. Platonist properties are not always had by more than one thing. The property of being Nelson Mandela is a property that only one thing has, so properties are not always multiply exemplifiable. Singleton trope property classes are the same. For the case where more than one thing has a property, we say that the property is multiply exemplifiable. Where there is more than one member in a trope property class,

[66] For our purposes, universalism will be Platonic, though Aristotle is also sometimes credited with holding a variety of universalism. He is also credited with holding a variety of the trope theory.

we say that it is multiply exemplifiable in that multiple things are members of the class by meeting the class membership condition characteristic of the class. Each member exemplifies the class in that it meets the class's characteristic class-membership condition in virtue of which the multiple members resemble one another.

If it is argued that the trope theory is not able to distinguish distinct properties with identical extensions, such as being triangular and being trilateral, the trope theorist can argue as follows. It can say that indeed it can distinguish between being triangular and being trilateral because having three angles and having three sides are distinct conditions on class membership, so they are distinct tropes.

Further, what the trope theorist does is explain how mental property tropes are distinct from physical property tropes, and yet the relata of causation, events, are psychophysically identical. That is, events with mental and physical properties are identical, and yet the property of being mental and the property of being physical are distinct. Now all that has been said in the Macdonalds' solution can be interpreted within a trope conception of reality, including events, and therefore there is no reason to suppose a trope-theoretic solution is deficient, unless there is an independent reason to reject the trope conception.

If we say that tropes are ways of being and that having a property is a way of being such that things that have it are bound by resemblance relations in virtue of that way of being that they have, we say that things share a property when they have ways of being bound by the resemblance relation which casts them into the class containing all things that are the resembling way they are. Then we can say that things have a particular property just in case they have a trope in virtue of which they fall into the class of relevantly resembling members- thus we say “the members share a property.” If the resembling relations are different for different tropes, then mental and physical properties are distinct in virtue of the fact that their characteristic resembling relations are distinct and can be expected to not share all their members. For example some neural tissue may bear a mental property at one time, but not at another, for example, when the organism to which it belonged is dead. This is not, however, to say that they cannot share some members. That is, the classes intersect, but probably neither wholly contains the other. Given this interpretation, we can say, as we did before, that mental properties of events are causally efficacious if and only if their instances are efficacious. Therefore, since properties can be co-instanced in a single event, given that some causally efficacious events have mental properties, mental properties are causally efficacious. The relevance issue would be dealt with in identical terms as the ones the Macdonalds use to deal with it, making use of the available literature on causation.

The Macdonalds object that the trope-theory’s ability to discern higher level from lower level properties of events does not secure the existence of mental properties. As the solubility example shows, the higher level property of being soluble is physical. At this point we should note that what was demanded by the Problem of Causal Exclusion was an account of consistency, or “reconciliation,” between certain theses, not a guarantee of the existence of mental properties. What we needed was a way to defend the causal relevance of mental properties, which presupposes their existence, from an independently motivated threat. As the Macdonalds themselves point out, the rejoinder is just that a trope theorist holds that it is particular higher level properties that are held to be mental, and not just any. This recognition should be more general than just applicable to trope functionalism. It is applicable to any particular way of being (trope) that is not the same as the physical way of being (like being mental), regardless of the intersection between physical and mental classes of events. That is to say, you can be a trope theorist that says that mental properties are not functional properties but are nevertheless not identical with physical properties. Further, the universalist conception may similarly be charged that since many higher-level properties are physical, as the solubility example shows, it does not ensure that mental properties exist. The universalist may plausibly respond as the trope-theorist does, saying that just because some higher-level properties are physical, that this does not rule out the unchallenged assumption that mental properties exist.

But the Macdonalds also say that we have reason *not* to call higher level properties, under the trope conception, “mental” “if all its members (i.e., tropes) are physical (Macdonald, C. and Macdonald, G. 2006, 553). However, not all possible members of the mental trope class are physical, since angels and immaterial souls would have mental properties. The intuition that this consideration has bite is tempered by the acknowledgement of the tricky and contentious issues surrounding the existence of possible (given that these are merely possible but not actual!), and the ontological commitments of “would have” in this context. However, it is worth noticing that this reason not to call higher-level properties “mental” applies to trope-theory only if it also applies to minimal physicalist universalism, the position the Macdonalds favor. This is because minimal physicalism already requires that all instances of mental properties are also physical, and therefore anyone who holds minimal physicalism, coupled or uncoupled with universalism, is given a reason to reject calling any higher-level property “mental.”

Perhaps the following analogy dissolves the considered worry that has now arisen for both camps. Just because every instance of the property of being square is a physical instance does not mean that the property of being square does not exist or is properly called “square,” or is not a different property from the property of being physical. *Mutatis mutandis* the same applies to the case of the mental and “mental.”[67]

The Macdonalds charge that:

> a more fundamental problem with the trope strategy remains, and that is that it simply doesn’t follow from the fact that P_i is an instance of both P1 and M1 that if P1 is causally relevant, so is M1. P_i will be an instance of any number of property-types, at least some of which (such as the property-type that is some heterogeneous class whose members bear only some weird kind of resemblance to one another) are plainly not causally relevant. Given this, we really have no reason to think the M1 is causally relevant (Macdonald, C. and Macdonald, G. 2006, 554).

Thus, the trope conception leaves an open door to the causal irrelevance of mental properties.

However, it seems reasonable that on any conception, trope or universalist, “weird” resemblance relations between things that “share” a property may not be causally relevant. Put another way, mental property tropes with weird resemblance relations will be causally irrelevant only if mental property universals specifying likewise weird resemblance relations are causally irrelevant. It may be countered that properties, on the universalist conception, are not constituted by their resemblance

[67] If it is claimed that this would not apply to the position of the Macdonalds as they give independent criteria for being mental, for example being rational, then the same applies to the trope theorist’s position. This is because, assuming they agree that the mark of the mental is being rational, this would select for beings that while physical, are also rational, even though not all members of the class of physical things are mental. Thanks to Graham Macdonald for posing the objection.

relations, and that because trope type properties are classes while properties on the universalist conception are not classes, they are not vulnerable to the same objection.[68] However, the universalist is committed to saying that because properties are multiply exemplifiable entities, every exemplification must resemble other exemplifications. This is because universalism is committed to saying that the numerically distinct exemplifications of the same property have a numerically identical quality and thus qualitatively identical things necessarily resemble one another as any one thing resembles itself. If multiple exemplifications of the same property are weirdly resembling with respect to that property, then having that property would likewise be causally irrelevant. It might be claimed that multiple property exemplifications with weird resemblance relations are impossible to begin with because identical qualities cannot be weirdly resembling any more than a thing can be weirdly self-identical. But then it is open for the trope theorist to apply the analogous reasoning to the trope type property case. The trope theorist can say that weird resemblances between tropes are likewise impossible to begin with, since such putative weirdly resembling tropes don't really resemble each other at all, any more than a money bank and river bank resemble one another in virtue of being weirdly resembling banks. They just are not banks in the same sense and therefore such weird resemblances are not real.[69]

[68] Thanks to Graham Macdonald for posing the objection.

[69] The universalist may, at this point, ask for clarification about what resemblance consists in since it is doing such significant work for the trope theorist in guarding from the possibility of weird resemblances. The trope theorist will in turn ask for a clarification of what a weird resemblance relation consists in, since it is the very existence of such a relation that would give force to the charge against the trope theory and it is unclear what that explanation would be, and if available, whether the universalist can recognize it as a universal. Perhaps, if it can, then the same problem applies to universalism; if it can't then universalism and trope theory remain in the trenches where they began: those of demanding and providing an account of a property.

But the Macdonalds charge that there is a deeper problem with trope theory. It is that it just is never able to account for the causal relevance of properties of events. This is because the trope theorist might claim that trope properties are classes, and events do not cause anything in virtue of being a member of class. The Macdonalds say that the trope theorist is not forced to defend MCR by construing mental properties as property-types, i.e., classes of tropes.

> She can maintain instead the *qua* problem can only arise at the level of individual tropes, and insist that if one accepts that mental properties are higher-level classes of physical tropes whose members fall into them in virtue of falling into lower-level classes, then there is no problem about mental causation or causal relevance at that level.[70] The grounds given for this might be that the items doing the work are not the classes but their members (the tropes); and no member does anything in virtue of its class membership. Indeed, it might be claimed, it makes no sense to say that a given trope caused (or figured in a cause of) *e* in virtue of being in one class or another (Macdonald, C. and Macdonald, G. 2006, 555-556).

However, the Macdonalds argue that this strategy, which casts putatively causally relevant properties as individual tropes rather than classes of tropes, ultimately results in merely answering the question of whether instances of mental properties are causally efficacious, while leaving the question of the causal relevance of the mental in a problematic state. The Macdonalds summarize the situation like this: “On either of the two readings of ‘property’ available, there is a problem of causal relevance. If the reading is ‘class of tropes’, then, for reasons already given, principle MCR is jettisoned. If, on the other hand, the reading is ‘trope’ the problem of causal relevance re-emerges, and the trope strategy does not solve it” (Macdonald, C. and Macdonald, G. 2006, 566-567).

[70] The Macdonalds attribute this is strategy to Heil and Robb (2003). See Heil (2003) also.

However, I would like to question the idea, independently of whether individual tropes are causally relevant, that “it makes no sense to say that a given trope caused (or figured in a cause of) *e* in virtue of being in one class or another” (Macdonald, C. and Macdonald, G. 2006, 555-556). The Macdonalds require that for properties of events to be causally relevant for an effect, there should be properties of the effect such that they figure, together with properties of the cause, in appropriate patterns. Now patterns are things that *repeat sequences of events*, so what most plausibly is able to do the job that the Macdonalds require are trope classes rather individual tropes. Does it make sense to say that a characteristic of an event is causally relevant, because it supposes that the event in question falls into a class? I think it does. For, it is an open move for the trope theorist to engage with the Macdonalds by accepting that questions of relevance should respond not only to efficacy but to there being a pattern subsuming the events in question. Such a trope theorist will say that indeed, the members of the class that satisfy the condition characteristic of that class behave in accordance with the pattern that make properties of the effect to be explained come about. As such, falling into a class can be a causally relevant property of an event, in virtue of which the event in question is said to cause an effect (with a particular character), for it signifies that the event in question belongs to a class of events that make events of the kind that the effect is come about. Sure, as the Macdonalds encourage us, the reading of “in virtue of” here must be, not in terms of efficacy which is not presently being questioned, but of relevance. It seems as if once the causal efficacy of a property is ascertained, the only question left to be answered about the causal relevance of that property is whether properties of the cause and effect fit into a pattern. If this is right, then it is open to a trope theorist who wants to address the Macdonalds' question of relevance, to answer in the same way that that they answer. Thus, at a general level, trope theory has conceptual resources that enable it to address questions of causal relevance to the same degree as the Macdondalds.

For the question of whether particularly mental properties are causally relevant, the same considerations the Macdonalds extrapolate from the literature on causation apply to a development of a trope-theoretic account of mental causal relevance. The trope theorist can legitimately claim that CF, as filled in by theories of causation such as the counterfactual and "two causes" models, serve to show how mental properties are causally relevant, just as the universalist can. The theories of causation are neutral with respect to universalism or trope-theory, and therefore are open for the universalist as well as the trope theorist to make use of them.

I take this to show that the trope theory is not a contender to the Macdonalds' solution, since each can deal with the issues particular to the Problem of Causal Exclusion just well as the other. Rather, the essentials of the solution are that (at least) two properties, one physical and one mental, are instanced in single causally efficacious event, such that the mental property is causally relevant, as evidenced by its ability to figure in the available literature on causation. Whether such properties are construed as the universalist or trope-theorist does is not relevant.

2.4.4 The Constitution View

The non-reductionist contender to the Macdonalds' solution, I think, is a theory that agrees with them that mental properties are distinct from physical properties, but that disagrees with them in that the higher level events are considered distinct from lower level events that realize them. Rather, lower level events constitute mental events. That is, a competitor would say that events that exhibit higher level properties are distinct from the events exhibiting lower level properties in virtue of which the former are realized. This account, as given in the literature in the philosophy of mind, is formulated under the assumption that *objects* are the causal relata. However, it is easily adapted to an account according to which the causal relata are events, as we have assumed. Let us begin with the object-based case for the non-identity of things that exhibit higher-level properties- which we are calling "higher-level object" and their counterparts "higher-level events," and then go on to extend the account to events.

Objects and their Constitution

Pereboom argues that entities at levels higher than the lowest exist and they do so because of the fact that they have different modal and temporal properties. The Ship of Theseus is not identical with any current token realization base since it has properties distinct from those of its actual realization base.[71]

> The ship of Theseus is not identical with its current token microphysical realization base, for it would have been the same token ship had the token microphysical realization been slightly different, and it will be the same ship when this microphysical realization in fact changes. The ship is in this sense *token multiply realizable* (Pereboom 2002b, 6).[72]

Because the ship survives changes in its constitution it cannot be identical with its constitution at any particular time. The ship is the same ship at different times, while its constitution is changed. For example, we can replace a plank, and we'd still have the same ship. We can gradually come to replace all the planks (doing it slowly enough, at least) and still have the same ship. So the ship cannot be identical to the collection of

[71] See Johnston (1992) for blocking objections of four-dimensionalism and other metaphysical frameworks to this non-identity claim.

[72] See also, Pereboom and Kornblith (1991).

planks that make it up at any particular time. This lesson is not uncontroversial, but Pereboom argues that it is an available, defensible view.

By Pereboom's lights the constitution relation is a realization relation. However, it should not be confused with the part-whole relation, though it is akin to it. The relevant notion of constitution is one that says that all the parts of the ship together at a time constitute, but are not identical to, the ship. The collection of all of a ship's parts, together and in the right organization realize the ship. But the collection of those parts in the right organization is not itself a part of the ship and therefore Pereboom's view does not imply a mereological account of realization. Further, no one would want to argue that a plank belonging to larger ship realizes the ship. At best it realizes a part of the ship. What is claimed is that the ship can and sometimes does have distinct constitutional bases, while remaining the same ship.

For constitution theorists, minds are, like ships, not identical with the microphysical objects that make them up. We can replace many quarks and electrons, perhaps all of them (if done in the right way), in a person's mind such that the mind will remain the same mind, while its realization base changes. Further, it is a fact that any person's microphysical constitutional base will change over time. The mind is, in this sense, token multiply realizable.

Why is this? It is because certain kinds of things, but more strongly, certain things, have properties that their realization base does not have. Such constitutive properties of higher-level things are demonstrated by their qualitative and systematic distinctness with respect to the lower-level things that make them up, by surviving differentially, changes to their modal and temporal conditions.

To take the case of having different temporal conditions on identity, an object's persistence conditions are the conditions under which it is identical to an object that exists in the future. These conditions on identity over time are different for the things that constitute an object from the conditions for the object they constitute to be the same object over time. The ship will survive the many losses of molecules and replacements that it will undergo in its lifetime, even when (supposing) the molecules are destroyed. Alternatively, the collection of molecules will survive, we may suppose, the destruction of the ship (supposing they are not destroyed). Similarly, the modal properties governing the identity of the ship and the identity of its parts are different, the supposed temporal scenario being one possibility where the ship and its parts wholly making it up come apart.

The same applies to minds and their constitution.

Extending the Constitution of Objects to the Constitution of Events

Now let us consider a particular event: The sinking of the ship. That token event has many smaller lower-level events composing it: the sinkings of the totality of its parts. Is the sum of sinkings identical with the ship's sinking? According to the constitutionalist No, since that same ship-sinking could have been composed of different part-sinkings- for example, if there were certain repairs done on it before leaving port, but which did not actually take place. In this sense, the event that is an exemplifying of the property of *being a sinking ship*, is distinct from events that compose it, *being sinking parts*. Since ship-sinkings are causally efficacious, and this very sinking of the ship is causally efficacious, just like with the Macdonald solution, we can refer back to theories of causation to justify its relevance.

Could we, however, question whether the constitutional account of higher level objects can be extended to events? At first glance, it looks like the same reasoning that justifies the object case applies equally to the event case. The lower level objects composing higher level objects just have different properties, so while they may coincide at a time, they do not share all their properties. In this sense, we have the same reason to accept the thesis that higher level events are constituted but not identical to lower level events as we do to accept the thesis that lower level objects constitute but are not identical to higher level objects. However, one of the ways in which higher level objects are seen to be distinct from lower level objects is that they, at least sometimes, maintain their identity past a time when their former constitution does not, or vice versa. In the event case, this seems to be pre-empted by the apparent fact that events are dated particulars and the lower level events that coincide with the higher level event have the same temporal coordinates essentially.

This apparent fact, the Constitution View may contend, is not really a fact though. The view might say that in fact the temporal coordinates of lower level constituting events do come apart from the coordinates of higher level constituted events, in that sometimes the sinking of the ship and the sinking of its parts coincide in time only partially. Such is the case of a ship that is being sunk and is fully submerged but has not reached the bottom. As the ship descends a powerful bomb detonates within it and its parts scatter, destroying the ship, but its parts survive and continue to sink. These parts' sinkings have temporal coordinates that outstrip the temporal coordinates of the ship's sinking. Thus, temporal properties for lower level and higher level events can differ as well, and the reason for holding a relation between lower level and higher level events to be constitution holds.

Now let us think about the mental case. How would the constitutional account go? A particular physical event realizes a particular mental event. We may suppose the physical event is the activation of a particular neural pathway x, which realizes a particular mental event, say, the having of a static visual image. The having of a visual image has different identity conditions than its realizing neural activation. The *having of a visual image*, we may suppose for this particular case, survives the *neural activation along a particular pathway*'s ceasing to be instanced (provided, for example, that it is replaced by the *activation along a near-by pathway* that does the same relevant job). The envisaged case is one in which the same static visual image had by a person, is constituted by, but not identical to, a neural event that person has, since she could have had the same visual image had some other neural event that did the same relevant job happened instead. In this case, different realizing events (activations along different particular pathways) are multiple realizations of the same static visual image. Only one of these is actualized, but the visual image has the property of possibly being realized by the other possible realization, while the activation along the actual pathway in question could not have been the same activation on a different pathway (the other possible realization of the visual image). Further, if hearing aids and other prosthetic devices are indicative of the reach of the multiple realizability of mental properties or events, then it is safe to say that the *having of a visual image* in question is realized by *silicon-based propagations* is possible. The constitution theorist says that the having of a static image is not identical to the having of a neural activation which realizes it, since the two have different properties. The relation between the lower level events and mental higher level events is constitution.

If we wish to understand this relation in terms of the instantiation of properties, as the PEA recommends, we can do it as follows (for simplicity we omit temporal reference). We can say that a brain has neural activations, the having of which are constitutive of this event. This event also has the characterizing property of being an instancing of neural activations. The person has a visual image, the having of which is a constitutive property of this event. This event also has a characterizing property: instancing the having of a visual image. The former instance constitutes but is not identical with the latter instance. Similarly, the former instancing constitutes but is not identical with the latter instancing.[73]

Defending the Constitution View from Hornsby's Reasons Against Mereology

Jennifer Hornsby (1985, 446) challenges the truth of the view that macro events are identical with fusions of micro events. Given that the Constitution View contends that mental events are constituted by, but not identical to, collections of micro events, particularly at least sometimes collections of neural events, it is worth seeing if the Constitution View can overcome the difficulty she sets out. For simplicity of exposition, "being constituted by" and its cognates, will imply being non-identical.

Hornsby says that mereological accounts of continuants will be committed to thesis A.

A. (x) (y) (E!z) (z is a fusion of x and y)

[73] This formulation is compatible with the requirement that the properties of the event have the property of being essential or constitutive properties of the thing in question as Kim's PEA requires, and what Baker's analysis of constitution (the one Pereboom most visibly endorses) requires. Baker's account is discussed later.

The problem with thesis A is that it "brings existential commitments far beyond any we ordinary recognize. For instance it guarantees that there is something composed from the Bodleian library and some carrot, and something made up from my copy of *The Structure of Appearance,* Goodman's left arm and your right leg" (Hornsby 1985, 447).

We might find an analogous principle in C-Fusion.

C-Fusion: (x) (y) (E!z) (z is constituted of a fusion of x and y)

C-Fusion clearly has the same problems as A. Hornsby claims that the mereologist will claim that A applies to material things, and that ordinary things just are among these. There are two usefully distinguishable components of A, namely E, which asserts the existence of fusions, and U which asserts their uniqueness.

E. (x) (y) (Ez) (z is a fusion of x and y)

U. (x) (y) (z) (w) (((z is a fusion of x and y) and w is a fusion of x and y) → (w=z)) (Hornsby 1985, 447).

She says that,

> The idea that two things can be in the same place at a time is a familiar, if recent, one. The contention is that we have to distinguish (e.g.) between a gold ring and the quantity of gold from which it is made, because the quantity of gold is something that may exist before the ring does and may go on existing after the ring is destroyed. A similar point can be made about almost any continuant and the molecules composing it at any particular time. And what leads one to distinguish two objects in cases like these is only an application of Leibniz's Law: x exists at t, y does not exist at t; so x is not the same as y (Hornsby 1985, 448).

Hornsby herself, then, seems to be committed to the Constitution View in the case of objects and (all) their parts. She also is committed to the view that "of course events sometimes stand to one another in part-whole relations" and this is just what one would expect of someone who holds the Constitution View about events as well.

Let us, however, see what analogous principles can be stated in terms of constitution rather than the relevant identity. We may note that given that she endorses the Constitution View, she should hope the analogous principles to the ones she attributes to the mereologist, i.e., the C-Principles (those prefixed by "C-") do not generate problems. To mimic E, we formulate C-Existence:

C-Existence: (x) (y) (Ez) (z is constituted by a fusion of x and y)

The Constitution View cannot honestly mimic U. We might try by formulating C-Uniqueness

C-Uniqueness: (x) (y) (z) (w) (((z is constituted by a fusion of x and y) and w is constituted of a fusion of x and y) → (w = z))

But the very essence of the Constitution View is that you can have more than one coincident entity constituted of fusions of distinct but coinciding entities. The quantity of gold constitutes a ring. The ring is constituted by a fusion of gold atoms. The fusion of gold atoms is itself constituted by a fusion of quarks and electrons. Both the ring and the fusion of gold atoms are constituted by a fusion of quarks and electrons. However, the Constitution View holds that the ring and the fusion of gold atoms are not identical.

What seems to survive as an unintuitive implication of the envisioned version of the Constitution View is the implication of C-Fusion that you have a thing that is constituted by but not identical with the fusion of the Bodleian library and some carrot, and something made up from my copy of *The Structure of Appearance,* Goodman's left arm and your right leg. The envisioned Constitution View will have the same response as the one Hornsby says the mereologist has: the response, that yes, these kinds of things are objects, so what?

However, Hornsby counters that, supposing the variables in C-Fusion range over events, such unrestricted fusions are implausible. This is because it is in the nature of events to cause and to be caused, and fusions of events need not be things that cause or are caused. “What on earth can we find to say about the causes and effects of a fusion of events whose parts occurred in 44 B.C., in 1066 and in 1984?” (Hornsby 1985, 450).

Hornsby says that a principle that might be used by the mereologist is C.

C. If (event c causes event e and f=c+d and (neither d occurs later than e nor d and e have common parts, nor part of e causes part of d) then f causes e (Hornsby 1985, 451).

Let us formulate the analogous principle for the envisioned Constitution View, C-Cause.

C-Cause: If event c causes event e and q is constituted by f (such that f=c+d) and (neither d occurs later than e nor, d and e have common parts, nor part of e causes part of d) then q causes e.

Hornsby says that based on her argument, those who hold typical counterfactual theories of causation “certainly” must reject C. However, she ends her argument with footnote 17, where she says that really her statements are only “a gesture toward an argument.” The reason she gives that the two views are incompatible is the following. Suppose c causes e, and therefore by counterfactualist account of causation “-c □→-e” is true ("□→" expresses the counterfactual-supporting conditional). If C is true, then "-(c+d) □→-e" is true. Hornsby recognizes that "this latter counterfactual is exactly what we shall expect to hold if the fusion of events had not occurred…" (Hornsby 1985, 451). She then says that such a defense ignores that "in any particular case of c causing e, it cannot be settled by sole reference to whether c actually caused e exactly which counterfactuals obtain. (It depends, for instance, on whether there was a fail-safe mechanism ensuring the production of e in the absence of c)." What can I say? Hornsby gives us a gesturing refutation at best, or a false advertisement at worst.

It is quite general, for any counterfactual that holds about c and e, that it will mention, either explicitly or implicitly, the causally relevant *properties* of c and e to determine which counterfactuals hold. The central counterfactual theory of causation is Lewis's, and according to his analysis, a counterfactual of the form "C □→ E" is true "just in case it takes less of a departure from actuality to make the antecedent true along with the consequent than to make the antecedent true without the consequent."[74] Such "departures from actuality" are going to be determined by the *properties* of the event in question. They are implicit, for example, when we say that the striking causes the lighting. They become explicit when we fill in causally relevant properties of the striking, such as that it was with a dry match, and of the lighting, such as that it was of the match. The counterfactual (putatively in accordance with making use of "sole reference" to the actual striking and lighting events) that "if the striking had not occurred, then the lighting would not have occurred" are determined to be true only if, in addition to reference to the events, such events have certain causally relevant

[74] I use Lewis (1973) theory, so as not to perhaps unfairly criticize Hornsby by alluding to the theory's development after her article was published in 1985.

properties. The striking could, logically speaking, *not have occurred*, but the match was lit by a lightning strike. In that case the subjunctive counterfactual putatively describing the causal interaction between the striking and the lighting, would not be true. The reason the scenario does not count, however, as a refutation of the considered counterfactual is that such a world would be very dissimilar to our world; and similarity of worlds, whatever your theory of properties, is measured by consideration of properties of the events at issue. [75] While events are the relata of causation, it is in virtue of their properties that they cause, and this fact is preserved under a counterfactual theory of causation. So it is quite generally true that which counterfactuals of a pair of causally related events hold cannot be determined by “mere reference” to those events, but there must also be reference to certain particularly causally relevant properties of those events. Thus it is unreasonable to demand of the mereologist, on the basis of the counterfactual theory of causation, that s/he must determine which counterfactuals hold of a pair of causally related events, by “sole reference” to those events. The account already presupposes, in addition, reference to properties of those events.

It is also highly controversial, at best, that there are fail-safe mechanisms, so it is highly controversial that determining the truth of causal counterfactuals depends on this. Fail-safe mechanisms would involve causal connections between events that cannot nomologically be interfered with. It is always hard to come up with candidates mechanisms of that kind, since it is hard to find any lawfully regulated mechanism that is not preconditioned by a *ceteris paribus* clause. Such clauses are put there because, if

[75] It is noteworthy that this is why scientific theory does not speak abstractly of “a causes b,” but rather gives a lot more information about a and about b.

you change those conditions, the posited mechanism would not be initiated or completed. So, if this worry, and it looks like a separate one disconnected from the counterfactual theory of causation, applies to the fusion case, it also applies to any uncontroversial cases of c causing e. That leaves the causing fusion in good company.

Hornsby, then, similarly fails to generate an apparent problem for C-Cause here.

Hornsby also tries to generate a problem for the mereologist on the basis of the regularity theory of causation. For the regularity theory,

> what distinguishes between a case where e causes f and a case where f follows on e temporally but not causally is that in the first case, but not the second, there is some significant (law-like or lawful) generalization that subsumes e and f. The problem about C for such a theorist is that it requires him to accept generalizations of an *ad hoc* and apparently trivial kind as playing a role which he has to insist only significant regularities play. Try to imagine some regularity that subsumes, on the one hand, the fusion of Caesar's death and my striking the match, and on the other hand, the match's lighting.[76] (Then) you must imagine as an interesting as possible (death+striking)/lighting regularity. But it is plain that *that* need not be law-like (Hornsby 1985, 452).

But Hornsby ignores that events fall into regularities in virtue of properties they have, so if the fusion of the striking and the death of Caesar inherits the property of involving a striking, then the fusion will have the lighting as an effect. That this "need not be law-like" is trivially true of any causal regularity, given that the laws of nature are contingent. Given that there is a regularity of the kind satisfying regularity theory between strikings and lightings, there is a regularity between (death+striking) and lighting.

One might counter that given that fusions inherit involving properties of their parts, that the fusion of being a particular death and a striking, will inherit involving the property of a particular death, and that it is plain that this property is not causally

[76] Hornsby is particularly thinking of Davidson (1967b).

related to lightings. The rejoinder is just that this property is causally related to lightings *only in that if* its instance is fused with an instance of a property that is causally related, it also is. By itself, without the fusion, it is not.

Further, some inherited properties of fusions will be causally relevant and some won't. The property inherited by the envisioned fusion of involving a particular death is not causally relevant for lightings; involving a striking is.

As long as this is clear, we explain why there is a distinction between the death and the striking, even if fused, with respect to the causation of the lighting.

Mutatis Mutandis the same consideration that Hornsby levels against C can be weighed against C-Cause, and *mutatis mutandis* the same response can be leveled to repel it. In fact, if this is the only avenue available for an advocate of C to respond, Hornsby should hope it works. This is because, given her acceptance of constitutional non-identity, she is committed to C-Cause, and C-Cause falls if C falls.

Back To the End of the Constitutional Account

Let us finish articulating the Constitution View. The Constitutional account would contend that the two events do not compete for causal efficacy and their properties do not compete for causal relevance. Given that one is constituted by the other at a time t, one is causally efficacious at t if the other one is causally efficacious at t. Given the resources from the literature on causation available to the Constitutionalist, the causal relevance of being mental is plausibly seen to be actual.

What would Pereboom's proposal say about EXC*? He contends that "stable tokens (given a certain level of complexity) often retain their identity over certain changes in the constitutions and configurations, and, significantly, they enjoy a certain resiliency in the production of their characteristic effect under these changes. So, for example, my decision to ring the doorbell can plausibly survive changes in its realizing microphysical state, and nature has likely endowed it with a resiliency for producing its characteristic effects under small changes" (Pereboom 2002b, 529). Pereboom wants to dispute that much weight should be given to a notion of the kind of dependence of a mental property on a physical property of an entity that abstracts away from its temporally and modally differing properties. He acknowledges that such a notion exists, but just contends that it is enough to secure non-identity that the mental and the physical differ in these properties, and that given the causal resiliency of the mental, in the context of varying microphysical bases, mental causes can be said to be irreducible to, or distinct from and independent from, the physical properties of the entity in question. Pereboom says that two causes constitutionally related create the possibility for having "two causal explanations for one event that do not exclude each other and at the same time do not reduce to a single explanation" (Pereboom 2002b, 505). As such, he rejects EXCL in that he thinks distinct and independent mental properties might well be causally relevant for their characteristic effect, even when their instances are constituted by (micro)physical ones.

It is not clear whether Pereboom has to reject EXC* in order to save the causal relevance of the mental. For surely, there is a sense in which the visual image is dependent on the existence on what constitutes it, and in that sense the event's having a mental causally relevant property toward some effect e, would be dependent on the microphysical events that constitute it. The dependency at issue would be supported by the constitution relation being asymmetrical, as Baker supposes (Pereboom 2002a).

Asymmetry about constitution says that, if a constitutes b, then b does not constitute a. However, Pereboom, while tentatively agreeing with Baker on her notion of constitution, does not always respect asymmetry. For example, notice the content of the parenthesis when he says:

> For if the token of a higher-level causal power is currently wholly constituted by a complex of microphysical causal powers, there are two sets of causal powers at work which are constituted from precisely the same stuff (supposing that the most basic microphysical entities are constituted of themselves), and in this sense we might say that they *coincide constitutionally* (Pereboom 2002b, 505).

The endorsement of the statement that microphysical entities are constituted of themselves ("a constitutes a") violates asymmetry, for microphysical entities are *rather* identical to themselves. Identity and constitution are supposed to function quite differently. Suppose that the microphysical entity a is identical to b. If so, if "a=b" then "b=a." But if there is supposed to be a constitution relation such that "b constitutes a" is true, it should not be the case that a constitutes b. "…constitutes…" should not function the way identity functions, for constitution is supposed to not allow symmetry, while identity does. Asymmetry is supposed to be feature of constitution that distinguishes it from identity. Pereboom seems to be allowing that "a constitutes b" follows from "b constitutes a" which is exactly the kind of inference presupposed by the idea that "a constitutes a."

We shall see that Baker's account of constitution does not account for asymmetry, either.

However, for the moment, if we take this statement to be a slip, and we recast it to his desired effect, that microphysical entities are identical to themselves, and do not constitute themselves, do constitute mental entities, and mental entities are identical to themselves, without constituting themselves nor microphysical entities, his theory is corrected, and we allow for the dependency evidenced by the asymmetry of the mental on the physical. Thus, Pereboom would reject EXC*.

2.5 Obstacles for the Proposed Solutions

Coming up are some obstacles for the success of both Macdonalds' solution by identity and Pereboom's solution by constitution.

In the way the Macdonalds formulate EXCL they point out how talk of properties as the apparent relata of causation can be used truthfully. They imply by the proposition embedded in EXCL, "a property, *P*, of a cause, *c*, is causally sufficient for an effect, e" that it is possibly true. But how can that be if supposedly it is only events, or instances of properties of events, that cause? As they point out, properties are said to be causally sufficient when their instances are instances that are causally sufficient. Given that this is so, the following objection can be formulated.

It is essential for the Macdonalds' Identity solution that the distinction between causal efficacy of events and the causal relevance of properties, where nomologicality is located, is respected. "The argument for non-reductive monism, if it works, works because of the extensionality of the causal relation, and the intensionality of nomologicality" (Macdonald, C. and Macdonald, G. 1995, 64). However, I would like to point to some elements destabilizing their conception. The case that there is a distinction between causal efficacy and causal relevance is brought out by an example of the Macdonalds. Suppose it is sufficient for a window to break that it is hit with a 5 pound or more rock at 10 mph, but not less. Actually, the window breaks because it is hit by a rock weighing 7 pounds at 10 mph. The property of the hitting that it is *of the window by a rock weighing over 2 pounds at 10 mph* is not *causally relevant* to the breaking, since there are rocks weighing between 2 and 5 (exclusively) pounds (at least possibly) hitting the window at 10 mph that do not result in the window shattering, and therefore this is not a causally relevant property. However, in this case, the *instance* of *being a hitting of the window by a rock weighing over 2 pounds* at 10 mph *is* causally sufficient for the breaking, since it is *a hitting with a rock that weighs 7 pounds at 10 mph,* since this is *a hitting with a rock that weighs more than 5 pounds at 10 mph*. So the account does not seem to count too many properties as causally relevant.

However, one may wonder whether the account permits too many properties of events to be causally sufficient, and whether by the same reasoning that attributes causal sufficiency too many properties are counted as causally relevant. Take the first wonder first.

2.5.1 Properties of Events Make Causal Contributions- But Surely Not All of Them

It seems that different properties of things make different causal contributions to the effects such things bring about. Suppose that the hitting with a rock that weighs 7 pounds at 10 mph that breaks the window is a hitting with a red rock. It seems that causal contributions of the sort illustrated by the redness component of the property of the hitting event are not matters strictly of relevance, but pertain to causal efficiency, which is what “causal sufficiency” is about. It seems correct, as the Macdonalds agree, that causation is an extensional metaphysical relation, and that causal contributions are precisely the sort of extensional metaphysical things that make effects, with their particular characters, come about. It seems that the redness that is a component *property* of the event in question makes no causal contribution to the breaking of the window, and it seems to me that this case is different from the case of being over 2 pounds in weight, which is an admittedly irrelevant but causally sufficient property of the event in question, *in the event in question* (if the distinction between causal relevance and causal efficacy is to be maintained).

It seems correct that we do not paint rocks in order to break windows with the resistances they have, and that we do not paint them not just because their colour is causally irrelevant to the breaking of the windows we might want to break, but because colour properties *contribute nothing simpliciter* to the breaking of windows, and in the case we are considering, more clearly this is true. You can vary the colour without varying the breaking in causal models at will, so long as you represent the different component properties of the property *being a hitting by a red rock weighing 7 pounds at 10 mph.* It may be argued that the causal models used to understand such scenarios are only sensitive to causal relevance, which is the domain of nomologicality and properties, which are only intensional, whereas the point I am making is supposed to be about events, the relata of causal sufficiency in the concrete world.

The Macdonalds use the literature on causal models to integrate mental causation (Macdonald, C. and Macdonald, G. 2006, Section 5). As they stand, however, it looks like such models are constructed with a blind eye to such a distinction between properties and events. Sometimes such models presuppose connections between properties, as when modeling the truth-conditions of statements like “taking birth-control pills causes thrombosis.” In such cases, what is relevant is that any event that has the property of being a taking a birth control pill by a woman raises the probability, along a causal route, of events having the property of being a having of thrombosis (Hitchcock 2001). Such models also work, however, on the basis of explicitly representing events, such as when modeling whether a particular worker’s severing of his hand causes it to function after getting fixed at the hospital, amongst other examples. A case in point where it seems that it is events rather than properties that are being modeled is in the case of actual causation. Actual causation is best modeled by component causes and the account that Hitchcock advocates of this is formulated in terms of events, rather than properties. The account of component causes he gives is: "Let c and e be two occurent events, and X and Y be variables representing alterations of c and e respectively. Then c is a cause of e iff there is an active causal route from X to Y in an appropriate causal model <V,E>."[77] It may also be noted that Hiddleston integrates an account of causal powers into such models. Causal powers are how things, associated with properties of events, get bound up in causal relations. Such causal powers correspond to different aspects of a cause-event and generate different aspects of an effect-event. According to these models, the colour properties of hittings by rocks do not have the causal power to break windows.

However, there is an option available to them for overcoming this difficulty, and it is to insist that such models are really sensitive only to causal relevance, for it is in terms of events' properties that the variables in such models are specified. One might take causally relevant properties to be properties that raise the probability of certain effects coming about. There seems to be a counterexample to this. The golfer intends to hit the golf-ball into the hole, but slices horribly. The ball ricochets off a conveniently situated tree and ends up in the hole. Here it looks like the golfer's slicing horribly decreased the probability of the ball going into the hole but was nevertheless a cause of the ball's going into the hole. However, the Macdonalds contend:

[77] Hiddleston (2005) explains this in section 1 on actual causes. <V,E> specifies variables representing putative causal relata and structural equations defining dependency relations between the values of the variables.

> Under some specification of properties, one involving forces acting on a body, what happens when a body in motion impacts on another body at a certain speed and direction, and so on, the probability of the ball's going into the hole *given that confluence of properties* may well be 1, or close to 1. And this combination of forces and resistances surely is involved when the golfer slices the ball. In general, what is 'accidental' is relative to a specification of properties. Given that different properties are exemplified in causally related events, the same effect can be accidental (or probability-decreasing) relative to one property-specification, and non-accidental (or probability-raising) relative to another (Macdonald, C. and Macdonald, G. 2006, 570).

Thus, we can say that the slicing horribly is a causally sufficient but, in the context of explaining the ball's going into the hole, irrelevant property of the cause; and that it is causally sufficient even if it actually decreased the probability of the ball's going into the hole. But now let's ask whether the property of *being a hitting by a red rock* is causally sufficient for the shattering. Well, if we adopt the scheme under which an event's exemplifying a certain property raises the probability of a given effect's occurrence, as an expression of the event's exemplifying of that property's causal contribution towards that effect, we understand the following. We understand that just as a property may be causally sufficient for the effect in question without being causally relevant, even when that property, as in the golfer's case makes a negative contribution, the hitting by a particularly *red* rock, which makes no contribution at all, may also be causally sufficient, while not causally relevant.

2.5.2 Too Many Causally Relevant Properties

One may also charge that too many properties are counted as being causally relevant. The contention is that from the supposition that an event has some causally relevant property, and the event is identical with an instance of all its properties, we get the result that all its properties are causally relevant; just as all its properties are causally sufficient. This is how the contention goes.

Take the event in question. It has the causally relevant property of being a hitting by a 7 pound rock at 10 mph, and it is identical with an instance of this property. The event's being identical with a hitting by a 7 pound rock at 10 mph is identical with a hitting by a rock weighing more than 2 pounds at 10 mph. One might think that since this hitting by a 7 pound rock at 10 mph is causally relevant, so is this hitting by a rock weighing more than 2 pounds at 10 mph (by the transitivity of identity). Therefore, we would be able to say that the event's being a hitting by a rock weighing more than 2 pounds at 10 mph is a causally relevant property. In that case, we lose our justification for the distinction between causal relevance and causal efficacy, and count too many properties as causally relevant.

One might try to block the reasoning before it even begins by saying that it is events that are the domain of causal sufficiency, whereas it is properties that are the domain of causal relevance. The argument, it might be thought, confuses properties and events, and correspondingly relevance and efficacy, and so begs the question. It might be thought that the argument's confusion, then, is based on the following scheme, which the Macdonalds reject. Properties are causally relevant if and only if their instances are causally relevant, and it is events that are their instances (supposing that it is events that are the relata of causation). Properties are causally relevant *only if* their instances are causally relevant because otherwise you could have causally relevant properties in a scenario where the only things that exist are entities in the purely Platonic realm where there is no causation, which would be absurd because being causally relevant presupposes making a causal contribution. If you don't have the proposed necessary condition on properties being causally relevant (that their instances be causally relevant), you could have causally relevant properties even if no causation took place, for example, in a scenario where the concrete world would not exist but the Platonic property realm did. (For completeness, but irrelevant to the argument at hand, properties are causally relevant *if* their instances are causally relevant because this bars the possibility that you might have all sorts of causally relevant instances of properties without having causally relevant properties.) Thus, since we talk in terms of properties in the context of causal relevance in order to talk of events, the relata of causation, it does no harm in this context to do the same for causal sufficiency, beyond taste, to talk of the causal sufficiency of *properties* as a way talking of the causal sufficiency of events in another manner of speech. Thus, in both cases, the truth-makers of statements attributing causal sufficiency and relevance are not, *ex hypothesi*, Platonic abstract entities outside of space-time, since the latter do not partake in the causal order of things, but are rather instances of such putative things, the causal relata, that is, events, that partake and make the statements true.

However, it is open for the Macdonalds to reject this scheme for begging substantive issues. They may well insist that they use “causal relevance” in the technical sense that it pertains to properties and indicate that it is different from causal efficacy in that while it shares the necessary condition on its application that its instances be causally efficacious, it also requires that the instance be an instance of a causal pattern that supports counterfactuals. Such a move would eliminate the unwanted possibility used to motivate the idea that it must be a necessary condition on the causal relevance properties that their instances be causally relevant. The unwanted possibility was that if rejected, there could be a world where there would be causally relevant properties without causation at all. Such an absurd world would only include abstract Platonic properties and no realm where causation took place, and yet there would be causally relevant properties. But according to the way the Macdonalds understand "causal relevance" such a possibility is also ruled out, as such a world would not contain causally efficacious instances of the putatively causally relevant properties. Thus, in such a world, the Macdonalds have the right verdict that there would not be causal relevance and therefore do not need the proposed alternative interpretation.

However, being causally relevant is a property of properties of things. Some properties have it and some don't. Properties that have causal relevance bear the relation that objects have to their properties: one has the other. The analysis the Macdonalds recommend of the first relation is that the relevant *having* relation is the exemplification relation; that is, objects instantiate, or exemplify, properties and are identical with instances of them. In case the things which have the properties are events, they will be identical with instancings of those properties. In the Macdonalds' solution, single instances have many properties, and the instances exemplify those properties. In their scheme, a single instance of those properties is the thing that has the properties. A "property is construed as a universal, and an instance of a property is not a trope of that universal, but a thing that has (instantiates, exemplifies) that property" (Macdonald, C. and Macdonald, G. 2006, 547). This is universalism[78]: an object that has a property is

[78] In formulating her account of exemplification or instantiation as an internal relation between a thing and the universal property it has, Macdonald (2005, 246) states examples of applications of this theory in places like "This instance of red, e.g. this red bird, is internally related to the universal, redness." At least for substances, universalism needs to be supplemented with a temporal index, since substances can have mutually exclusive properties at different times. We do not want to say that if the box is an instance of being totally monochromatically red at one time and an instance of being totally monochromatically green at another time, that the instance of being totally monochromatically red is,

identical with an instance of that property. In causal contexts, such objects are events and are instancings of the properties they have. They say, as an example, "Consider a red, square box. It has the properties of being red and being square. It also has the property of being colored. It is (identical with) an instance of each property that it has" (Macdonald, C. and Macdonald, G. 2006, 563). This thesis allows them to say that all properties of a thing are causally sufficient for the effect the thing has. This makes all properties of an event candidates for being causally relevant.

How might we say, then, that instances of properties of properties are had by objects? Well, instances of properties are identical with the things that have the properties. By the same standard, instances of properties of properties are identical with the things that have them. Take the hitting of the window. It has the causally relevant property of being a hitting with a rock weighing 7 pounds at 10 mph. By universalism, it is identical with an instance of being such a hitting. This property of the event itself is an instance of being causally relevant. The hitting's instancing of being by a rock weighing more than 7 pounds is also identical with an instancing of being a hitting by a rock weighing more than 2 pounds at 10 mph. By the transitivity of identity, the hitting's being by a rock weighing more than 2 pounds and at 10 mph is causally relevant.[79]

Schematically, if Fa is causally relevant and Fa is identical with Ga, then Ga is causally relevant. The idea that Fa is identical with Ga, in this case, is licensed by

in the sense of identity, an instance of being totally monochromatically green. Let us maintain the temporal index implicit for the main discussion.

[79] This would seem to fit in with Cynthia Macdonald's (1989, 161) Principle of the Nomological Character of Causally Relevant *Instances* of Properties since it makes clear that the relevant properties of properties have identical instances. "Associated with any event type, such as 'shoots', is a property, such as that of being a shooting. This property will be instanced by any event (i.e., is identical with) a token of that type. Indeed, the sense in which any token event 'has' such a property consists in the fact that it instances that property."

universalism- since it says that Fa and Ga are identical with a. This is because a thing's having a property is its instancing (in case it is events) of that property and an instancing of a property just is the thing that has it. The contention is that for any Fx, if Fx is causally relevant and any G, distinct from F, such that Gx, then Gx is causally relevant. This is the contention that too many properties are cast as causally relevant. Or:

1. **$(P)(x)\ (Px \rightarrow ((x = P_i)\ \&\ (P_i = Px))$** (Universalism, see the Macdonalds' conception above). Read: *For any property P and any event x, x's having P implies that x is identical to an instancing of P and the instancing of P is identical to x's having P.*
2. **Ga** (Premise Intro) Read: *a's having G*
3. **Fa** (Premise Intro) Read: *a's having F*
4. **$(a = G_i)\ \&\ (G_i = Ga)$** (By application of 1 (Universalism) to 2) Read: *a is identical to an instancing of G and the instancing of G is identical to a's having G*
5. **a = Ga** (Application of Conjunction Elimination to 4 and subsequent application of Leibniz's Law) Read: *a is identical to a's having G*
6. **$(a = F_i)\ \&\ (F_i = Fa)$** (Application of 1 (Universalism) to 3) Read: *a is identical to an instancing of F and the instancing of F is identical to a's having F*
7. **a = Fa** (Application of Conjunction Elimination to 6 and subsequent application of Leibniz's Law) Read: *a is identical with a's having F*
8. **Ga = Fa** (Application of Leibniz's Law to 5 and 7) Read: *a's having G is identical with a's having F*
9. **(Causally Relevant) Ga** (Premise Intro) Read: *a's having G is causally relevant*

10. (Causal Relevant) Fa (Application of Leibniz's Law to 8 and 9) Read: *a's having F is causally relevant*

"F," "G," and "a" were chosen arbitrarily, and therefore, apply to any event with at least two properties at least one of which is causally relevant. The result is that given a causally relevant property of an event, any other property of that event will be causally relevant.

Perhaps the way the Macdonalds can block this apparent result is by assimilating causal relevance to belief. It is generally accepted that in belief-contexts co-referring terms cannot be substituted (via Leibniz's Law) guaranteeing the preservation truth-value. Lois's believing that Superman can fly does not imply that she believes Clark Kent can fly, even though Clark Kent and Superman are the same person. However, their given conditions for causal relevance, that its instance be causally effective and fit into pattern, do not immediately bring the assimilation out. Moreover, their window shattering example is not justified on the basis of imperfect information, as the belief-context case likely would. More work must be done for them to overcome the difficulty specified here.

2.5.3 The Identity Solution Does Not Allow for the Existence of the Relation of Non-Identical Constitution

It is another problem of the Macdonalds' solution that it does not allow for the existence of constitution without identity. Non-Reductive Monism implies the existence of contingent identities, since the Macdonalds countenance the PEA of events, whose component Existence Condition and Identity Condition on events distinguishes between constitutive and contingent properties of events. Given that constitutive properties of events are the properties that make those events the events they are, while they could have or fail to have their contingent properties, the events' identity is contingent. For example, spinning up when a quark spins up is considered a constitutive property of the event in question. This spin might well be my favorite event, but this is merely a contingent property of the event. The event that is a spin up and the event that is my favorite event are contingently identical. For the event is contingently identical with my favorite event, a contingent property of it.

Now take the case of the sinking of the ship. The Macdonalds will say that the sinking of the totality of its parts is identical with the sinking of the ship. For any putatively plausible relation of constitutional non-identity between things, they are committed to there being an identity, of a contingent kind.

The same thing goes for mental causes. They will be contingently identical with physical events.

In conversation, Cynthia Macdonald has told me that she does not reject the possibility of non-identically constituting objects; it is just that this does not hold for physical and mental events. Perhaps there is a sensible position that says that while some objects are constitutionally related, others are contingently identical. However, if the two views do not compete it is hard to see why they have been posed as alternative accounts of the relation between the physical and mental, such that we have more reason to endorse one view rather than the other.

2.5.4 The Constitution Solution Does Not Allow for the Existence of Contingent Identities

On the other hand the Constitution View has problems of its own. In particular, it seems to not allow for contingent identities. By its own reasoning, for example, we might have to count a thing's being red by being scarlet red as pertaining to non-identical objects, since a thing's being scarlet red has different modal and (sometimes) temporal properties from its being red. Supposing that the thing in question changes from scarlet red to crimson red, its being a red object survives, while its being a *scarlet* red object is destroyed. If such differences in temporal and modal properties are sufficient to say that ships and their parts are related by constitutional non-identity, we need a reason not to be committed, by the same reasoning, to say that the red and scarlet red objects are distinct. This reasoning applies for all multiply realizable properties, objects, and events. However, we are not given a reason to block this extension, even though Pereboom thinks the determinate/determinable relation is "more intimate than constitutional coincidence" (Pereboom 2002b: fn.15). But it seems at first glance like this extension is not only inevitable, given the reasoning he has used, but is misplaced. It seems like an instance of being scarlet red just is (in the sense of identity) an instance of being red.

It is plausible that in cases, not just pertaining to the determinate/determinable relation, the constitutional non-identity is the wrong relation. It seems that when x has the property of being Ed and of being the winner of the race, Ed is identical with the winner of the race, and that while x's being Ed and being the winner of the race carry different temporal and modal properties onto the thing that has them,[80] we do not say that for this reason the thing that has them must be distinct, even though being the winner is multiply realizable, as Pereboom agrees.

These cases are not easily blocked off as Pereboom does not delineate what sort of relation the constitution relation is and where it can or cannot hold. He does however

[80] For temporal distinctness: Ed existed prior when no one had won the race and therefore there was no winner. For modal distinctness: Someone else other than Ed could have been the winner, and yet, by definition, whoever is the winner is the winner.

think that Baker's general notion has a serious chance of being right (Pereboom 2002a), so it is worth seeing what she says on this matter. Baker believes that one of the virtues of her view is that there need not be contingent identities, or any other "faux identities" (Baker 2002, 35).[81]

At least to begin, it may be that we don't need contingent identities and we would want the Constitution View to tell us how such putative identities are to be truly understood. In that case, let us put Baker's formal apparatus to work. Baker thinks that:

> constitution is a relation between concrete individuals. Each concrete individual is fundamentally a member of exactly one *primary kind*. By definition, any concrete individual has its primary kind membership essentially, so that a concrete individual x'*s* ceasing to be a member of this kind entails that x ceases to exist. For example, (*Goliath*'s) primary kind is *statue*, (Lumpl's) primary kind is *(Lump of clay)*. Suppose that *x* and *y* are concrete individuals; *F** designates the property of *having F as one's primary-kind*; F and G are not the same kind; individual x has F* and individual y has G*; and *D* designates *G-favorable circumstances* -- "the milieu required for something to be a G" (39-42). Then, (C) x constitutes y at t $=_{df}$ (a) x and y are spatially coincident at t; and (b) x is in D at t; and (c) it is necessary that: Az[(F*z at t & z is in D at t) implies Eu(G*u at t & u is spatially coincident with z at t)]; and (d) it is possible that: (x exists at t & ~ Ew[G*w at t & w is spatially coincident with x at t]); and (e) if y is immaterial, then x is also immaterial (Pereboom 2002a, pageless website version).[82]

[81] In the case of brains and minds, at least for the most part, even by Gibbard's (1975) scheme, brains and minds are not counted as identical, for most people's brains outlast their minds- so the two come apart in fact, and this is indicative of non-identity. In the case of actual minds, then, they are not identical with brains. However, the contingent identity of Gibbard's theory might be complicated to incorporate a mind-brain identity thesis by adding temporally relativizing notions of identity like "x is contingently identical to y at time t but not identical at t'," and therefore "Ed (a mind) is identical to Brainy (a brain) at time t but not after Ed's death, when Brainy continues to exist." However, see also Lewis (1976, 1991). In these he draws an analogy for personal identity without complete temporal coincidence from roads. Roads might coincide only for a portion, but while they coincide, they are just one road (illustratively, we need not cross two roads to cross roads while they coincide).

[82] I have substituted Baker's and Pereboom's "piece of marble" for "lump of clay," "Piece" for "Lumpl" and "David" for "Goliath." This way we fix on the example, rather than switch between examples with no significant difference. The cases Baker considers as individuals are objects (substances) and the account may need reworking to apply to events. Events are, however, individuals. As Cynthia Macdonald has pointed out, at least some events are changes such that an object goes from having a static property to having another, exclusive property, in some time. However, for the time being we will stick with individuals as objects.

This formalization subsumes the case of Lumpl and Goliath. According to this case, as famously put forth by Allan Gibbard, Lumpl, a piece of clay, and Goliath, a statue of Goliath the man, come into existence at the same time and cease to exist at the same time. Two hunks of clay, adequately shaped, are put together at a time, and both Lumpl and Goliath come into existence. At a later time, Lumpl and Goliath are destroyed by the reverse mechanism. Lumpl and Goliath share all their actual "obvious properties," colour, shape, and coincide spatially at every instant of their existence. As things stand, Lumpl and Goliath share all their fundamental particles or point instants. Thus, Gibbard wants to claim Lumpl and Goliath are identical. However, they might not have been. If he had taken the statue and squeezed it into a ball, Lumpl would have survived, while Goliath might not. Thus, Lumpl and Goliath are contingently identical (Gibbard 1975, 191-192).

Baker on the other hand says that Lumpl and Goliath are two things, constitutionally related in the way she defines. Even if they don't come apart Lumpl and Goliath have different properties. Lumpl constitutes Goliath, since Lumpl and Goliath at t, for any time of their existence, spatially coincide at t (condition a). Lumpl is at any time of its existence in Goliath-favorable circumstances without interferences (condition b), Lumpl has the property of having as its primary kind, *piece of clay*, while Goliath has as its primary property *statue* (condition c). Further, if any object whose primary kind is *piece of clay* that is in Goliath-favorable circumstances without interferences, then there is an object whose primary kind property is *statue* and is spatially coincident with the piece of clay at t. It is possible that Lumpl exist at t, while there is no object that has as its primary kind property *statue* spatially coincident with

Lumpl at t (condition d). If Goliath is immaterial so is Lumpl (condition e). The definition is designed to say just how Lumpl and Goliath are related and bring out why they are different. According to Baker, primary kind properties are essential properties of the things that have them, and therefore can be expressed by modal operators. Thus, essential properties are modal properties, and Baker believes they belong to objects in a way that justifies the application of Leibniz's Law. Since Lumpl and Goliath have different essential properties, they must be different objects.

The problem is that any property of an object can specify an object with that property as essential to it, that is, without which the object would not be the thing it is. If this is right, the theory is too open in specifying too many objects to which the constitutional relation applies- and therefore gives us no insight, for many properties, to how objects having them constitute others. This is because it results in, for any case, very many objects having any of their properties as primary kinds, constituting other objects of distinct primary properties. Suppose that Goliath was painted scarlet red. In Baker's scheme, Goliath and the *scarlet red thing*, call it Scarlet, share the property of being scarlet red. Yet, Goliath non-identically constitutes Scarlet, since they satisfy Baker's Constitution definition. True, we don't normally name scarlet red things just as we don't normally name pieces of clay. But that does not mean they are not nameable. In such a case, for some t, Goliath and Scarlet at t are such that Goliath and Scarlet are spatially coincident at t (condition a). Goliath has been painted scarlet red the day before t, and we may harmlessly suppose that there have been no interferences with this state of affairs, and therefore Goliath is in the favorable circumstances required to be a scarlet red thing (condition b). It is necessary that anything of whose primary kind is *statue* (Baker’s choice for Goliath’s primary kind) and has been painted scarlet red the day before t, without interference in between, implies there is something whose primary kind is *scarlet red thing* at t (my choice for Scarlet’s primary kind), which is spatially coincident with the statue at t (condition c). Since Goliath might not have been painted scarlet red, it is possible that Goliath exists at t and there is nothing having as its primary kind property the property of being scarlet red that is spatially coincident with Goliath at t (condition d). Condition e is trivially satisfied in this case, as neither is immaterial. Pereboom’s preferred version of condition d is put like this:

> Either it is possible that (x exists at t and it is not the case that y is spatially coincident with x at t), or it is possible that (y exists at t and it is not the case that x is spatially coincident with y at t) (Pereboom 2002a, pageless website version).

Since Goliath might have not been painted scarlet red Scarlet would not have been spatially coincident with Goliath at t, so Pereboom’s version is satisfied by the example. However, there may be some qualm about Scarlet’s primary kind being *scarlet red object*. Although Baker would prefer to not use the term “substance” because of associations with substance dualism, which she rejects, she does think that she is a substance pluralist if forced to use the term, and it may be that someone is suspicious that the proposed primary kind for Scarlet is not a substance-kind. However, the kind at issue definitionally includes object-hood, a paradigm of substance-hood, such that Scarlet has the primary kind property of being a *scarlet red object*. I must confess that I cannot think of any property *of an object* that cannot be specified as a primary kind property, without saying anything more or less than that the object belongs to the primary kind, of having a given property, like being scarlet red. This may at first be put thus:

FREE: Necessarily, Ax (Fx → F*x)

However, some properties of a thing are derivative and therefore non-primary kind properties, in Baker's scheme. In her scheme, Lumpl would have the property of being a statue but only derivatively, and so contingently so. So, that Goliath is scarlet red does not imply, with necessity, that Goliath must be scarlet red. Rather it must be this:

FREE 1: Necessarily, Ax (Fx → Ey (F*y))

FREE 1 is an interesting thesis. It says that for each and every property a thing has, there is a thing that has that property as a primary kind property.

Further, another thesis extracted from Baker's conception is:

By FREE 1, to every distinct property of an object, there corresponds a primary kind property of the object. For every primary kind property of an object, there is an object with a distinct primary kind that constitutes it, according to Baker's and Pereboom's definition. Given that contingent identities specify their objects in terms of distinct properties (flanking the identity sign), to each object with a property, any putative contingent identity will be transformable into a constitutional relation.

It is a corollary of this statement that according to Baker's scheme, not only does Lumpl constitute Goliath but Goliath constitutes Lump.

This is particularly embarrassing for the Constitution View given that constitution is supposed to be an asymmetric relation. If a constitutes b then b does not constitute a. Giving asymmetry up, Baker says, is tantamount to giving up the project of constitutional non-identity. But this is the present result.

Baker says that Lumpl and Goliath are two things, constitutionally related in the way she defines. Even if they don't come apart Lumpl and Goliath have different properties. Lumpl constitutes Goliath, since Lumpl and Goliath at t, for any time of their existence, coincide at t (condition a). Goliath is at any time of its existence in Lumpl-favorable circumstances without interferences (condition b), Lumpl has the property of having as its primary kind, *piece of clay*, while Goliath has as its primary property *statue* (condition c). Further, if any object whose primary kind is *statue* that is in Lumpl-favorable circumstances without interferences, then there is an object whose primary kind property is *piece of clay* and is spatially coincident with the statue at t. It is possible that Goliath exist at t, while there is no object that has as its primary kind property piece of clay spatially coincident with Goliath at t (condition d). If Goliath is immaterial so is Lumpl (condition e).

It is a corollary that for any two properties of an object, there will be two corresponding objects constitutionally related and never related by identity.

2.5.5 Primary Properties in the Metaphysics of Constitution Are Uninformative

Given that what a primary kind is, is *all* that is required of a thing to be in order to be the thing it is -"the milieu required for something to be a G"- when we test our intuition about what could be the F*'s and G*'s, we are lead to specifying too uninformative and maximally specific properties of x to be the F*'s or G*'s. In my view, the project would not be immediately in trouble if it said that say, a set of molecules non-identically constitutes Lumpl, which in turn non-identically constitutes Goliath, and so forth. The Constitution View in this case could well get some points by not having the consequence that many objects are identical to one object.[83] The problem begins to appear when we begin to wonder what *the* primary kind property of Goliath is. For example, Goliath, in Baker's choice, is presumably countenanced to be *statue*. However, we may begin to be suspicious of this, given that we have to pick one property to be its all-important primary kind property and the statue may survive its morphing into a statue of a shark, as a statue, while not surviving as Goliath, the statue of a human. Then we might say the statue is of the primary kind *human statue*. But then of course, we might morph the Goliath statue to be a replica of David, thereby destroying the human statue of Goliath and bringing into existence the human statue of David. The response would be to specify more precisely what property is had by Goliath without which it would not be Goliath. It looks like the road will lead to the trivializing conclusion that that the primary kind of Goliath is being itself, and similarly for any other individual. This would amount to saying that Goliath has the property of being Goliath as its primary kind, since no other property is sufficient to uniquely pick it out through all alternative required properties.

2.5.6 The Debate between Contingent Identity and Constitution

What is Baker to say about Gibbard's case of contingent identity? In Gibbard's case, Lumpl, a piece of clay, and Goliath, the statue, come into existence at the same time, and are destroyed at the same time. Gibbard thinks that this is a case of contingent identity, while it may be true that Lumpl and Goliath have different modal properties. However, Gibbard thinks that modal contexts are to be differentially treated from

[83] Which is not to say that some theorists don't have resources available to overcome the problem of saying how many objects are equal to one. Theorists committed to the idea that objects are identical with the set of objects composing them are committed to such a view. What is noteworthy, though, of the constitutional analysis, is its straighforwardness with which the apparent problem is resolved.

contexts in which we talk about concrete individuals. In cases where we are in modal contexts we are, by contrast, talking about concepts (which cannot be contingently identical), like the concept of a statue or a lump of clay.[84] He argues this on the basis that since modal properties are not properties of concrete objects independently of how they are designated, and modal properties of objects are always, if putatively had, had only relative to a way a thing is designated, modal properties are not properties of concrete objects.

> A property, if it is to be a property, must apply or not to a thing independently of how the thing is designated…Expressions constructed with modal operators… simply do not give properties of concrete things, such as statues and pieces of clay. Lumpl, for instance, is the same thing as Goliath: it is a clay statue of the infant Goliath which I put together and then broke. Necessary identity to Lumpl, though, is not a property which that thing has or lacks, for it makes no sense to ask whether that thing, as such, is necessarily identical with Lumpl (Gibbard 1975, 201-202).

Thus, concrete objects do not have modal properties, Gibbard (1975, 201) claims. This allows him to say that the differing modal properties of Lumpl and Goliath are not real properties that can be used, via Leibniz's Law, to generate the conclusion that the things that have them are distinct. Rather, when modal contexts are created variables range over concepts rather than concrete objects. Thus, if Gibbard is right in following Carnap on the interpretation of what modal statements quantify over and what they don't quantify over, Goliath and Lumpl cannot be straightforwardly said to have different modal properties. Baker's argument in this article relies on the premise that Goliath has the property of being essentially a statue, whereas Lumpl does not have this property, and therefore by Leibniz's Law, Goliath and Lumpl are distinct (Baker 1997, 601). Of course, statements that express essential properties are expressed via the

[84] Gibbard (1975) fills this out in detail.

necessity operator of modal logic, and in this sense essential properties are modal properties.

Baker does not think that Gibbard is correct on the question of whether modal properties can be attributed to individuals. She thinks that when a doctor says, as he pulls a bullet out of a soldier's flesh, "that could have killed you," that the doctor may say a truth about the individual object that is the bullet. She says that predicates which attribute dispositions, probabilities, and causal powers to concrete things apply to concrete things only if modal expressions apply to those things independently of how they are specified (Baker 1997, 606). Given, thus, that such predicates do apply to concrete things, modal predicates must apply to those things independently of how they are specified. This, however, on my view is not correctly argued, since it seems plausible that the appearance of "that" in the expression by the doctor serves contextually to mean a specification like "the object I just pulled out of your flesh" and Gibbard is free to argue that the doctor may well express a modal truth insofar as the designation "the object I just pulled out of your flesh" allows it. And in terms of dispositional properties, there is no argument to suggest here that it is not necessarily specified *qua* salt, or some other property, that certain compounds have the dispositional property of being soluble; that it is specified *qua* salt that a given probability belongs to certain aggregates of molecules for dissolving in certain circumstances; and that it is *qua* salt that certain aggregates of molecules have the causal powers to dissolve in water. Further, Baker unfortunately does not engage with Gibbard's system so as to be able to claim that he does not have an adequate treatment of statements which appear in everyday and scientific contexts employing such modality-involving notions and therefore does not say why his system is inadequate. For disposition talk, for example, Gibbard gives a re-interpretation that allows us to maintain it- though not *straightforwardly*, to use Gibbard's (1975, Section IX) terminology.

In my view, Baker is, however, fundamentally correct in saying that Lumpl and Goliath have different modal properties. Admittedly, Gibbard's system is able to incorporate any statements we want to say. However, modal properties are given a radically new interpretation at the price of losing *de re* modal properties of concrete things. Modal properties are said to apply to individual concepts but not concrete individuals, and modal statements that seem to be about individuals must be reinterpreted to be about concepts in order to be applicable. It seems to me that this semantical move complicates our view of modal statements by interpreting them differently, rather than the same as, statements ascribing non-modal properties; and it does this to include the rather distasteful and implausible view that concrete objects do not straightforwardly have *de re* modal properties, while concepts do.

It seems to me that the same reason that moves Gibbard to saying that individual concrete things do not have *de re* modal properties applies as well *mutatis mutandis* to individual concepts. Gibbard is an essentialist about concepts and not about concrete things, where essentialism applies to "the fulfillment of conditions by objects…apart from special ways of specifying them…"[85] "Essentialism for a class of entities U…is the claim that for any entity e in U and any condition * which e fulfills, the question of whether e necessarily fulfills * has a definite answer apart from the way e is specified" (Gibbard 1975, 207). He thinks that it is Goliath designated as a statue,

[85] Gibbard (1975, 161) quotes W.O. Quine (1961).

that can yield that it is essentially a statue, and “Goliath” by itself does not do that. But he thinks that if we must designate Goliath as a statue, or under any other sortal, in order to say whether Goliath is necessarily a statue, Goliath’s necessarily being a statue is not a property of it.

But now take the case of concepts. Gibbard thinks that it is important to be an essentialist about concepts, since he takes it that the Carnapian system he adopts implies it.

Furthermore, it seems no less reasonable than the case of Lumpl and Goliath’s contingent identity, if such a relation is to be given at least a chance, that in the following scenario the individual brain structure and the individual concept of person P are contingently identical. This is supposed to be a scenario, in the first instance, where concepts are things we, who live and think wholly embedded within nature, have and are things that play a causal role in thought and behaviour patterns. Suppose the concept had by P is about Goliath, call it “C-Goliath,” and the brain-structure that is coincident with it “B-Structure.” That is, on the supposition that Lumpl and Goliath are identical, we allow for the same reasons, that C-Goliath and B-Structure in P are identical. In the envisaged scenario, C-Goliath and B-Structure begin and end at the same time. If Gibbard is correct that concepts have straightforward modal properties, then C-Goliath has straightforward modal properties (that is, independent of the way it is designated), while B-Structure does not have straightforward modal properties. In that case, then, C-Goliath has straightforward modal properties and B-Structure does not, and therefore they cannot be identical. That is, if Gibbard is correct that concrete individuals do not have straightforward modal properties, but are said to have them only *qua* a certain specification, then because B-Structure in P has modal properties only *qua* certain specification (a neural structure), while C-Goliath in P has them independent of the way it is designated, C-Goliath in P and B-Structure in P have different properties and are distinct.

On the other hand, if C-Goliath and B-Structure are identical, and C-Goliath has straightforward modal properties, so does B-Structure. If straightforward modal properties apply to B-Structure, there is no reason to deny that modal properties apply to many other concrete objects.

2.6 A Dissolution of the Problem of Mental Causation

We have seen that there are two versions of the Problem of Causal Exclusion. In the display I gave of Kim's formulation, it was emphasized that the mental properties and their instances are dependent on their physical realization bases. I will therefore call this the "Dependency Version." In the Macdonalds' formulation emphasis was laid on the independence of the mental properties from their realization bases. I will therefore call this the "Autonomy Version." Further, these versions can be analyzed in terms of properties, and in terms of their instances. What we end up with is a two by two model, like this:

	The Dependency Version	**The Autonomy Version**
Causal Efficacy of Instances of Mental Properties	**Causally Excluded by the Physical?**	**Causally Excluded by the Physical?**
Causal Relevance of Mental Properties	**Causally Excluded by the Physical?**	**Causally Excluded by the Physical?**

This way of analyzing the situation ensures consistency in the use of properties and instances, and the version of the problem we are concerned with. It also does not require us to use EXC*, as it is covered by the Dependency Version.

I will show a clear and perfectly generalizable case of supervenient causation where there is no competition between supervenient and subvenient elements, and look at how the two, or four, versions of the Problem of Causal Exclusion look when this conception is in place. The Dependency Version has been stated in terms of the efficacy of instances and the Autonomy Version in terms of the causal relevance of properties, though both versions can each be thoroughly stated in terms of the causal relevance of properties and the causal efficacy of instances.[86] My solution contrasts with the deeply metaphysical, putatively competing, approaches that have been taken to the problem. I say that my solution is a *dis*solution because I use no heavy-duty metaphysics; my solution is compatible every metaphysics standing in the philosophical market today as providing a solution. Also, it is a *dis*solution because I argue that either the mental is included within the causally sufficient physical factors that bring about physical effects, or the problem relies on false principles. Either way, there is no problem left standing.

2.6.1 Supervenient Causation for the Dependency Version

There is at least one important kind of supervenience that permits supervenient phenomena to be causal, or not epiphenomenal. In fact, this kind of supervenience is of the kind that at least many mental phenomena belong, according to a prominent dualist, David Chalmers. This kind of supervenience is logical supervenience. Logical supervenience implies the strongest sort of dependency as it has the strongest modal force. The supervenience base is accompanied by its supervenient elements in the most amount of possible situations, i.e. all of them, where the supervenience base is held

[86] Of course, this presupposes that instances can supervene on other instances, as a constitution theory such as Derk Pereboom's (2002a; 2002b) may require, if constitution is a supervenience relation. However, if instances of supervenient and subvenient properties are taken to be identical (even if the properties are distinct), as Cynthia Macdonald and Graham Macdonald (2006) take them to be, then supervenience may not be taken to hold between instances. This is because supervenience is normally taken to be an asymmetric relation, whereas identity is symmetrical. However, if that is the case the exclusion of one instance by the other is precluded by the identity, and we need not worry about the case. I include it only to incorporate token, or instance, constitutionalists, like Pereboom. While I do address the formulation in terms of supervenience as a relation between properties, also I give it a showing in terms of instances for completeness.

fixed. It also allows for the independence that is attributed to supervenient mental properties via their multiple realizability.

To recap, “B-properties supervene logically on A-properties if no two logically possible situations are identical with respect to their A-properties but distinct with respect to their B-properties” (Chalmers 1996b, 35). Logically supervenient properties are such that they cannot vary, given the status of their supervenience base. From an informational point of view, if you have adequate definitions of supervenient properties, you can in principle read them off their supervenience bases.

With the view that we wish to understand Kim’s argument in terms of this analysis, and Kim’s argument, as I presented it, is formed in terms of instances of properties, let us analyze logical supervenience in terms of instances of properties. Kim’s argument can also be reformed to be framed in terms of properties, and then the above analysis will be the useful one. The Autonomy Version is done in terms of properties, below.

For property-instances, “Instance(s) of property B supervene(s) logically on instance(s) of property A if no two logically possible situations are identical with respect to their A-instance(s) but distinct with respect to their B-instance(s).” The intended interpretation of this analysis is premised on the assumption that just as properties can be the same across possible situations, so can instances of properties. There are several contending analyses for such cases, and the reader can interpret them according to her general approach. No new problems will arise that are not already present in the interpretation of sameness claims about properties and individuals across possible situations.

Now, suppose that the realization base of an event's instancing of the supervenient property of turning red is its instancing of the distinct property of turning scarlet red. Suppose that this is an event in a traffic light, which causes drivers to stop. One may notice that the light's turning red is dependent on its turning scarlet red in that its turning scarlet red, determines in this case, its turning red. One may also notice, that the light's turning red is also, in another way, independent of the light's turning scarlet red; for red is multiply realizable. The light could have turned crimson red, for example, and still have turned red and, importantly, caused the stopping. Could the event's being a turning red (its supervenient property) be causally excluded by its being a turning scarlet red (its realization base)?

Let us substitute into 3' thus:

> 3'' Stopping is instantiated on this occasion: (a) because an instance of the traffic lights turning red caused an instance of a stopping (as an instance of supervenient causation), or (b) because an instance of the traffic lights turning scarlet red, the traffic lights turning red's instance's supervenience base, causes the stopping to be instantiated.

One would think that it is obvious that there is no choice to be forced between (a) and (b), and the corresponding instances as causally efficacious, and properties they posit as causally relevant. It seems to me that this case is projectible to supervenient mental properties, whether they be logically or naturally so; since both are cases of supervenience where the supervenient properties are dependent on their bases, in that they are in any particular case, determined by them; and also independent of their bases, in that they could have been realized by an alternative realization. These are the features in virtue of which the example is projectible, to other cases of logical and natural supervenience.[87]

Recall *Closure*: If a physical event has a cause that occurs at t, it has a sufficient physical cause that occurs at t. The stopping in our example is caused by the stoplight's turning scarlet red. This turning scarlet red essentially involves turning red. When we say that there is a sufficient physical cause for the causation of the stopping, by *Closure* this must be the total set of factors that causally contributed to the stopping. For of course, there are many actual and possible circumstances under which stoplights' turning scarlet red fail to make cars stop. So, the turning scarlet red "*by itself*" is not sufficient for the causation of the stopping. Rather, the sufficient cause is the turning scarlet red plus many other conditions, just as striking a match is sufficient for a lighting just in case it is compounded with a combustible gas around it, that it is against a rough surface, with a dry match, etc. Thus, for example, we will have to include within this set, the attention of the driver, visibility conditions of the stoplight, the direction the car was heading towards, etc., including the turning red (see below).

But now consider again *Exclusion*: No single event can have more than one sufficient cause occurring at any given time- unless it is a genuine case of causal overdetermination. Could this principle exclude the turning red from being a causal contribution for the causation of the stopping? Well suppose we took the set of conditions that are the sufficient cause for the stopping, and skimmed off the turning red. If we do this, we skim off the turning scarlet red. Then, *ex hypothesi,* the remaining set of conditions would not be sufficient for the stopping. The supervenient property-

[87] The reader will notice that Edwards' dictum (below) is supposed to apply to both logical and natural supervenience.

instance of being a turning red is thus seen for the causally efficacious contribution to the stopping it makes. Thus, either supervenient properties and their instances are allowed to be part of the sufficient physical cause, or either *Closure* or *Exclusion* is false.

Either way, the supervenient instance of turning red does not compete for causal efficacy with the instance of being scarlet red for the causation of the stopping, even though one is superveniently dependent on the other; and the supervenient property of turning red does not compete with turning scarlet red for causal relevance, even though one is superveniently dependent on the other.

The same applies in the natural supervenience case, as in this case, if supervenient properties are skimmed off, then given that natural supervenience is the kind that holds in virtue of the contingently given laws of nature, you must have changed the laws of nature for the supervenient base to obtain, and not the supervenient properties. Some of these cases will involve making the supervenience base insufficient for the effect in question. For example, if you vary the K constant to half its value in the pV=KT as it regulates a particular mole of gas in a volume at a temperature (a fixed supervenience base), you decrease the pressure it exerts by half, and therefore makes it insufficient, for example, to cause the barometric reading it may have caused under the natural scenario. Again, either supervenient properties and their instances are allowed to be part of the sufficient physical cause, or either *Closure* or *Exclusion* is false.

2.6.2 Supervenient Causation for the Autonomy Version

Now let’s look at the principles of the Autonomy Version. Take CLOS: If a physical event or phenomenon has any cause, it has a sufficient physical cause, whose physical properties are causally sufficient for its effect- unless it is a genuine case of overdetermination; and take EXCL: If a property, *P*, of a cause, *c*, is causally sufficient for an effect, *e*, then no other property, *Q*, distinct from and independent of *P*, is causally relevant for *e*.

Again, either the sufficient physical cause with its causally sufficient physical properties includes the supervenient property of turning red or it doesn't. If it does then there is no reason to think it must be excluded by properties of the rest of the physical cause. If it does not, then we take out the stoplight's turning red, and by implication, we take out the stoplight's turning scarlet red, the putative physical subvenient causally relevant property with which turning red competes. But now, *in that scenario,* the stopping is not caused, and either CLOS or EXCL is false. So either supervenient properties and instances of them are allowed be part of the sufficient cause, or either CLOS or EXCL is false.

2.6.3 Some Corollaries and Edwards' Dictum

A. At least some supervenient instances of properties, whether dependent or independent in the above sense, do not compete for causal efficacy with instances of their subvenient properties. That is, one does not exclude the other.

B. At least some supervenient properties, whether dependent or independent in the above sense, do not compete for causal relevance with their subvenient properties. That is, one does not exclude the other.[88]

[88] The reader may note that the traffic light's turning scarlet red and turning red is, by Chalmers' analysis of "natural supervenience," a case of natural supervenience, since no naturally possible

One might try to bite the bullet and say that somehow the turning scarlet red but not the turning red causes the stopping. Perhaps Kim can say there is competition based on Edwards' dictum, which says that "There is a tension between vertical determination and horizontal causation. In fact, vertical determination excludes horizontal causation" (Kim 2005, 36). Kim gives two reasons for it but I find no good justification for accepting this principle. First he confounds a question of causal ancestry with a question of "reductive explanation" as explained above (2.2.1) in the case of giving an explanation for the existence of a Carbon atom by asking whether a statue's being yellow at a time is explained by its subvenient properties at that time or by its being yellow before. (Remember also that this is supposed to justify somehow the reasoning in the argument in the first place, not use the reasoning.) Secondly, Kim says that somehow the mental is like a mirror image, which never depends for its existence on the existence of prior mental images, but rather on being caused at each time. One may note that this is not a case of supervenience in the first place, so there is no strict vertical determination. If there is an analogy with simultaneous causation (perhaps as simultaneous as it gets), all it proves is that some cases of apparent simultaneous causation are really epiphenomenal. There are equally plausible cases of genuine simultaneous causation. A central case of simultaneous causation is where the point of drill is caused to turn by the turn of its base, even though they turn at the same time (or at least, as simultaneous as such cases get). However, it would be crazy to think that the turning of the base of the drill competes with the turning of the point of the drill for the

situation in which the traffic light turns scarlet red, does it not turn red. It is not considered a case of natural supervenience with the analysis recommended here, as it should be.

causation of a hole being drilled into the wall, and that the turning of base excludes the turning of the point.

The dictum then seems to be really based on an abstract fear that it might somehow be true, and if true, wouldn't it be terrible? In my view it does not have a proper justification. Those who bite the bullet bite it do it at the cost of all the things that (at least for some of us) require mental causation (given that it requires supervenient causation), like agency (given that it requires rationality), the mental itself (given that we should not believe in things that make no causal difference to the world), and cognitive science (given that it seeks to discover the mental causal mechanisms that generate behaviour). With respect to the theoretical value, it has the cost of saying that the stoplight's turning (scarlet) red causes the stopping without its turning red- a contradiction. What is the benefit for this cost?

2.6.4 A Dissolution of the Dependency Version

Now we have enough to dissolve the Problem of Causal Exclusion in its two forms and return the recognition of the mental as a causal phenomenon. Let us look into it bit by bit. Suppose, again, we have an instance of mental to mental causation. This would be a situation represented like this:

3. An instance of a mental property M1, occurring at t1, causes an instance of the mental property M2 occurring at t2.

If we are to be physicalists, then there must be an important sense in which M1 and M2 are not only mental properties, but also physical properties. Since physicalism is the doctrine that everything that exists is physical, if mental properties exist, they must be physical. The way in which properties are physical is that instances of them are exhibited in space-time and governed by the laws of nature. While this may work with trope conceptions of properties, some may want to hold a Platonist view of properties, and for the Platonist *properties as such* can never be physical. Since Platonists conceive of properties as outside of space-time and the causal flux of the universe, properties for them could never enter into causal interactions, whether "physical" or "mental." However, they should allow that there is a significant sense in which at least some properties, like having particular mass, are physical, and that having such properties enables the things that have them to enter into causal interactions of particular sorts. The way to do this is to remember (2.2.7) the following:

Physical Sufficiency Condition on Properties: properties are physical if their instances are physical.

Physical Sufficiency Condition on Instances: instances of properties are physical if they are in space-time and regulated by the laws of nature.

As applied to the particular case in question we might note that M1 and M2 can be said to be physical properties, if their instances are in space-time and governed by the laws of nature. Given physicalism, instances of M1 and M2 must be physical, as well as mental. By the Physical Sufficiency Condition on Properties, while M1 and M2 are mental properties, they are also physical properties, just like being neurons, having a particular mass, or being a square, are physical properties of some entities, etc.

Kim says that it follows from our analysis of supervenience that:

4. The M2 instance has a physical supervenience base P2 instance, at t2.

Now if the instance of M1 is going to cause an instance of M2, it must cause its supervenience base, an instance of P2. Thus, a supervenient mental property instance is supposed to cause a subvenient physical property instance. This shows that: "Under the mind-body supervenience assumption, mental-to-mental causation implies, or presupposes, mental-to-physical causation" (Kim 1998, 43). So, the question now becomes whether we can make sense of mental to physical causation. From what we have discovered about the physical, there is no mystery about mental-to-physical causation, since the "mental" in "mental-to-physical causation" is itself physical- as much as being as a stoplight's turning red, a toy being square, a cell being a neuron, and so on, are physical. As such, the interaction can also be described as "physical-to-physical," and this is not a kind of causation supposed to be at risk here- supposing that causation itself is not at risk.

On the other hand, we can argue that mental properties are not part of the sufficient physical cause of M2 or P2 because of *Closure* and *Exclusion.* However, given that the mental supervenes on the rest of the physical, just as turning red supervenes on turning scarlet red and the rest of multiple realizations of it; just as skimming off turning red skims off turning scarlet red (and any other multiple realization of red), if you skim off the mental from the rest of the physical you skim off its multiple realizations. If you skim off the mental's multiple realizations you get rid of the causally efficacious instances and the causally relevant properties that make up the causal factors supposedly sufficient for the effect in question. For example, suppose you got rid of desires to turn on lamps in the world. Then, you get rid of all the subvenient multiple realizations of desires to turn on lamps. These realizations are what Kim says competes with desires, for it is a given in the debate, that they surely cause the turning on of the lamps that are turned on putatively because we desire to do so. But now you have made the sufficient cause an insufficient cause, and you are going get a lot less turning off of lamps. What is brought into effect in the world changes if you take out the mind. So either the mental is included under "sufficient physical cause" or it is not. If it is, then there is no exclusion of the mental, and it can maintain its causal status. If it is not, then either *Closure* or *Exclusion* is false. Either one of them is false because if you take out the mental, you take out the effects of the realization of its cause, falsifying *Closure. Closure* gives the impression that the sequence of events in world would be the same with or without the mind; which is patently not true. Or, if the mental is not part of the sufficient physical cause, then it must be false that an event cannot have more than one sufficient cause unless it is a case of genuine determination, falsifying *Exclusion.*

Thus,

	The Dependency Version
Causal Efficacy of Instances of Mental Properties	**Not Causally Excluded by the Physical**
Causal Relevance of Mental Properties	**Not Causally Excluded by the Physical**

2.6.5 A Dissolution of the Autonomy Version

The Autonomy Version can be dissolved by noticing that just like the firing of c-fibers, the having of mental images and other mental phenomena, are physical things; or otherwise CLOS or EXCL are false. In the first horn, when we say that (PCR) physical properties of physical events are causally relevant to the physical effects those events bring about, we do not exclude mental properties from being such physical properties any more than we exclude being a c-fibre firing, an electromagnetic attraction, or a moving square.

Further, if we think that there will always be other properties on which the mental supervenes, we can point to the case of the red traffic light to indicate how a distinct, and independent property (in the sense required, i.e., of multiple realizability), does not compete for causal relevance with its supervenience base property. If we don't include the mental as part of the sufficient physical cause, and the mental property as a contributor to the physical property P, then either CLOS or EXCL is not true, for the same reasons given above. Namely, just as you take out the causally relevant property of turning scarlet red by taking out the turning red, you at least sometimes take out the physical causally relevant properties by taking out the mental ones.

The Macdonalds do not consider instances of properties to be causally relevant, since causal relevance is not about instances but about properties. Since, however, a property cannot be causally relevant without its instances being causally efficacious, if there is no threat in the Autonomy Version for mental properties to be excluded from being causally relevant, there is no threat about physical instances excluding mental instances from being causally efficacious.

Thus,

	The Autonomy Version
Causal Efficacy of Instances of Mental Properties	**Not Causally Excluded by the Physical**
Causal Relevance of Mental Properties	**Not Causally Excluded by the Physical**

2.6.6 The End of the Problem?

The result is:

	The Dependency Version	**The Autonomy Version**
Causal Efficacy of Instances of Mental Properties	**Not Causally Excluded by the Physical**	**Not Causally Excluded by the Physical**
Causal Relevance of Mental Properties	**Not Causally Excluded by the Physical**	**Not Causally Excluded by the Physical**

The Problem of Causal Exclusion is an argument that will probably continue to attract interest. I have attempted to give the problem yet another go (or two or four), reformulating it in a consistent way, both in terms of properties and in terms of their instances, for the two versions it is broken up into. There have been deeply metaphysical proposals designed to deal with the Problem of Causal Exclusion. My formulations aim to capture the problem these proposals aims to answer. My solution is a dissolution because it is compatible with all of them, and yet it solves the problem.[89] Further, it says that either the mental is allowed to be part of the "sufficient physical cause," in which case, there is no Problem of Causal Exclusion; or either the closure or exclusion principles are false, in which case, again, there is no Problem of Causal Exclusion. In my view, the mental lives without this metaphysical guillotine hanging over it.

[89] I have in mind the central metaphysical solutions of Kim, the Macdonalds, Pereboom, as well as the trope-theoretical solutions.

The Road Taken

Let us sum up our epistemic journey by grouping some our conclusions. Our starting point was the perception of certain obstacles in holding Mental Realism. One came from the philosophy of science, while the other came from the philosophy of mind. We amended a standard statement of Scientific Realism to allow for the existence of the bearers of truth-value (beliefs, statements, theories, etc.) and the existence of minds. To do this we got the following statements of Scientific Realism, with an application to mental reality:

1. The Metaphysical Thesis: The things in nature, sometimes mental and sometimes not, we aim to refer to with our scientific theories, make our theories true or false.

Its application to mind is:

1M. The Metaphysical Thesis about Mind: The mental things we aim to refer to with our psychological theories make those theories true or false.

2. The Semantic Thesis: Scientific theories are truth-conditioned descriptions of nature, including portions of nature that are not objects of observation.

2M. The Semantic Thesis about Mind: Scientific theories of mind are truth-conditioned descriptions of mental portions of nature, including mental portions of nature that are not objects of observation.

The Epistemic Thesis could remain the same, while we dealt with a particular potential threat coming from Searle's problems with computational cognitive science. With an application to mind, it is:

3M. The Epistemic Thesis about Mind: Current psychological theory is well-confirmed and approximately true of the world. So, entities posited by them, or at any rate, entities very similar to those posited, do inhabit the world.

In the process of defending and refining these theses, we covered a lot of ground. I provided new criticisms to positivist irrealist philosophy of science, as well as to van Fraassen's Constructive Empiricism. We also covered a new argument immanent in Searle's work against computational cognitive science. We saw how that argument is not sound, because it assumes a structuralist philosophy of computation, which is wrong as Max Newman had shown Russell's structuralist philosophy of physics is. This structuralist philosophy presumes that computational properties of entities in nature are not intrinsic, empirically discoverable, nor causal. I argued that there is no reason to believe the theory of computation is committed to structuralism. We defended the **Useless Cardinal Thesis** which states that if it is true that any aggregate has any structure compatible with the same number of parts, then any object has any structure. We also made some proposals for the theory of computation and questioned the internal consistency of the theoretical underpinnings of Searle's criticisms.

In the second section we started by formulating the Problem of Causal Exclusion and made some amendments to it to make it more compelling. Notably, by distinguishing causal from "reductive" explanation. We saw how it applies to some prominent positions in the philosophy of mind, such as functionalism and anomalous monism. I defended a notion of the physical that would overcome Chomsky's challenge to physicalism, and as well as the problems encountered by Poland's, Pettit's, Papineau's, and Dowell's notions. I advocated a conception of the physical according to which anything that is wholly in space-time and regulated by the laws of nature is physical, and physicalism is just the thesis that everything has these qualities. I called my proposal ***Unpacking Physicalism***. We noted, however, that there was still work to be done to dispel the Problem of Causal Exclusion, particularly the work that addresses what supervenience does in the argument. In this line, we also made a suggestion for how to understand natural supervenience, one that overcomes the problem that Charmers' analysis has. The problem was that Charmers' notion would count some cases of logical supervenience as cases of natural supervenience, when it is clear that they are the former but not the latter.

Then, I suggested that Kim believes the following inconsistent ideas about higher-level causation throughout his long term development of the Problem of Causal Exclusion:

> ***Real*****:** That things with non-basic properties exist, with their own causal powers.
>
> ***Epiphenomenal*****:** But that since irreducibly non-basic properties of things are wholly causally dependant on basic properties of things, irreducibly non-basic properties of things are causally epiphenomenal with respect to basic properties of things.
>
> ***Reduction*****:** Non-basic properties of things are identical with basic properties of things that have them.

We then went into the most prominent metaphysical solutions to the Problem of Causal Exclusion, Non-Reductive Monism and the Identity Theory, and the Constitution View. We noted problems for them, as well as some of their theoretical motivations. We saw that while Non-Reductive Monism, if it does not imply Platonism about properties, is compatible with Trope theory, and has the resources to overcome one of the objections that if its principles are upheld every property of an event is causally relevant. However, as far as I can see, there is another objection to the same effect that is successful. We then noted that the metaphysics of constitution is usually cast as an alternative to contingent identity, and we noted a feature of the relata of constitutionally related objects and objects related by contingent identity (**FREE 1**): that for any property of an object, there is an object that has that property as its primary kind property. The "object" in the sequent is either contingently identical or constitutionally related to the object in the antecedent. Further, I contended that the Constitution View does not have the resources to preserve the asymmetry essential to its theory. Further, if Non-Reductive Monism and the Constitution View are seen as competitors, then they fail to recognize that some objects are constitutionally related (without identity), while others are related by contingent identity (without constitution).

We further noted that primary properties in the metaphysics of constitution are uninformative; and that the theoretical underpinnings of contingent identity theory are unable to give a consistent modal characterization of a concept a human may have. Concrete things have no *de re* essential properties, while concept do. But then a concept a human may have, both has and does not have *de re* essentialist properties.

In the previous chapter we noted that in fact we can systematize the Problem of Causal Exclusion by dividing it into a two by two model, cross-cutting the Dependency and Autonomy Versions by the Causal Relevance of Properties and Causal Efficacy of Instances. Here I argued that the problem, in all its variants, is dissolved because either we include mental causes in the causally sufficient factors that generate physical effects or the problem relies on false principles. Either way, there is no problem left standing.

Mental Realism lives on.

References

Armstrong, D.M. 1970. The nature of mind. In *Readings in philosophy of psychology* vol. 1, ed. Ned Block, 191-199. Cambridge: Harvard University Press, 1980.

Baker, Lynn Rudder. 2002. On making things up: Constitution and its critics. *Philosophical Topics* 30, 31-52.

Bickle, John. 2006. Multiple realizability. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/multiple-realizability/

Block, Ned. 1980. Introduction: What is functionalism? In *Readings in philosophy of psychology* vol. 1, ed. Ned Block, 171-184. Cambridge: Harvard University Press, 1980.

Boyd, Richard 1997. Realism, approximate truth, and philosophical method. In *The philosophy of science*, ed. David Papineau, 215-255. Oxford: Oxford University Press, 2001.

2006. Natural kinds as social artifacts: Implications for realism and reference. *University of Canterbury Research Seminar.*

Burton, Harold. 2003. Visual cortex activity in early and late blind people. *The Journal of Neuroscience* 23, 4005-4011.

Chalmers, David. 1996a. Does a rock implement every finite-state automaton? *Synthese* 108, 309-333.

1996b. *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.

Chomsky, Noam. 2000. *New horizons in the study of language and mind.* Cambridge: Cambridge University Press.

Churchland, Paul. 1982. The anti-realist epistemology of Bas van Fraassen's *The scientific image*. *Pacific Philosophical Quarterly* 63, 226-235.

Copeland, Jack. 1996. What is computation? *Synthese* 108, 335-359.

1997. Vague identity and fuzzy logic. *The Journal of Philosophy* 94, 514-534.

2000. Narrow versus wide mechanism. *Journal of Philosophy* 97, 1-32.

Cummins, Robert. 1975. Functional analysis. In *Readings in philosophy of psychology* vol. 1, ed. Ned Block, 185-190. Cambridge: Harvard University Press, 1980,.

1991. *Meaning and mental representation.* Cambridge: MIT Press.

Davidson, Donald. 1967. Causal relations. *Journal of Philosophy* 64, 691-703.

1970. Mental events. In *Readings in philosophy of psychology* vol. 1, ed. Ned Block, 107-119. Cambridge: Harvard University Press, 1980.

1993. Thinking causes. In *Mental causation,* eds. Alfred Mele and John Heil, 3-17. Oxford: Oxford University Press, 1993.

Dowell, Janis. 2006. The physical: empirical, not metaphysical. *Philosophical Studies* 131, 25-60.

Edwards, Jonathan. 1758. Doctrines of original sin defended. In *Jonathan Edwards,* eds. C.H. Faust and T.H. Johnson. New York: American Books Company, 1935.

Egan, Frances 2003. Naturalistic inquiry: Where does mental representation fit in? In *Chomsky and his critics*, eds. Louise Antony and Norbert Hornstein, 84-104. Oxford: Blackwell, 2003.

Evans, Gareth. 1973. The causal theory of names. *Aristotelian Society Supplementary Volume* 47, 187-208.

Fodor, Jerry. 1992. *A theory of content and other essays.* Cambridge: MIT Press.

Friedman, Michael. 1975. Physicalism and the indeterminacy of translation. *Nous* 9, 353-374.

Friedman, Michael and William Demopoulos. 1985. Bertrand Russell's *The analysis of matter*: Its historical context and contemporary importance. *Philosophy of Science* 52, 621-639.

Gibbard, Allan. 1975. Contingent identity. *Journal of Philosophical Logic* 4, 187-221.

Grice, Paul. 1975. Logic and conversation. In *The logic of grammar,* eds. Donald Davidson and Gilbert Harman, 64-75. California: Dickenson, 1975.

Heil, John. 2003. *From an ontological point of view.* Oxford: Clarendon Press.

Heil, John and David Robb. 2003. Mental properties. *American Philosophical Quarterly* 40, 175-196.

Hempel, Carl. 1958. The theoretician's dilemma: A study in the logic of theory construction. In *Concepts, theories and the mind-body problem,* eds. Herbert Feigl, Michael Scriven, and Grover Maxwell. Minnesota Studies in the

Philosophy of Science 2, 37-98. Minneapolis: University of Minnesota Press, 1972.

1977. The logical analysis of psychology. In *Readings in the philosophy of psychology* vol. 1, ed. Ned Block, 14-23. Cambridge: Harvard University Press, 1980.

Hiddleston, Eric. 2005. Causal powers. *The British Journal for the Philosophy of Science* 56, 27-59.

Hitchcock, Christopher. 2001. A tale of two effects. *The Philosophical Review* 110, 361-396.

Hornsby, Jennifer. 1985. Physicalism, events and part-whole relations. In *Actions and events: Perspectives on the philosophy of Donald Davidson*, eds. Ernest LePore and Brian McLaughlin, 444-460. Oxford: Basil Blackwell, 1985.

Jackson, Frank. 1994. Finding the mind in the natural world. In *Philosophy and the cognitive sciences*, eds. Roberto Casati, Barry Smith, and Graham White. Vienna: Holder Pichler-Tempsky, 1994.

2002. From reduction to type-type identity. *Philosophy and Phenomenological Research* 65, 644-647.

Jackson, Frank and Philip Pettit. 1990. Program explanation: A general perspective. *Analysis* 50, 107-117.

Johnston, Mark. 1992. Constitution is not identity. *Mind* 101, 89-106.

Kim, Jaegwon. 1984. Supervenient and epiphenomenal causation. In his *Supervenience and mind*, 1993, 92-108. Cambridge: Cambridge University Press.

1989a. Supervenience as a philosophical concept. In his *Supervenience and mind*, 131-160. Cambridge: Cambridge University Press, 1993.

1989b. The myth of non-reductive materialism. In his *Supervenience and mind*, 265-284. Cambridge: Cambridge University Press, 1993.

1989c. Mechanism, purpose, and explanatory exclusion. In his *Supervenience and mind*, 237-264. Cambridge: Cambridge University Press, 1993.

1993a. The non-reductivist’s troubles with mental causation. In his *Supervenience and mind*, 336-357. Cambridge: Cambridge University Press, 1993.

1993b. Postscripts on mental causation. In his *Supervenience and mind*, 358-368. Cambridge: Cambridge University Press, 1993.

1993c. Multiple realization and the metaphysics of reduction. In his *Supervenience and mind*, 309-335. Cambridge: Cambridge University Press, 1993.

1998. *Mind in a physical world.* Cambridge: MIT Press.

2002. Précis of *Mind in a physical world*. *Philosophy and Phenomenological Research* 65, 640-643.

2003. Blocking causal drainage and other maintenance chores to do with mental causation. *Philosophy and Phenomenological Research* 67, 151-176.

2005. *Physicalism, or something near enough.* Princeton: Princeton University Press.

Kornblith, Hilary. 1980. Referring to artifacts. *Philosophical Review* 89, 109-114.

1989. The unattainability of coherence. In *The current state of the coherence theory*, ed. John Bender, 207-214. Dordrecht: Kluwer, 1989.

1992a. Our native inferential tendencies. In *Readings in philosophy and cognitive science,* ed. Alvin Goldman, 69-94. Cambridge: MIT Press, 1993.

1992b. The laws of thought. *Philosophy and Phenomenological Research* 52, 895-911.

2002. *Knowledge and its place in nature.* Oxford: Claredon Press.

2007. How to refer to artifacts. In *Creations of the mind*, eds. Eric Margolis and Stephen Lawrence, 138-149. Oxford: Oxford University Press, 2007.

Kripke, Saul. 1972. *Naming and necessity.* Cambridge: Harvard University Press.

Kujala, Teija; Palva, Matias; Salonen, Oiji; Alku, Paavo; Huotilainen, Minna; Jarvinen, Antti and Naatanen, Risto. 2004. The role of blind humans' visual cortex in auditory change detection. *Neuroscience Letters* 379, 127-131.

Lewis, David. 1970. How to define theoretical terms. *Journal of Philosophy* 67, 427-446.

1972. Psychophysical and theoretical identifications. In *Readings in philosophy of psychology* vol. 1, ed. Ned Block, 207-215. Cambridge, Harvard University Press, 1980.

1973. Causation. *Journal of Philosophy* 70, 556-567.

1976. Survival and identity. In *Self and identity*, eds. Daniel Kolak and Raymond Martin, 273-285. New York: Macmillan, 1991.

1980. Mad pain and martian pain. In *Readings in philosophy of psychology* vol. 1, ed. Ned Block, 216-222. Cambridge, Harvard University Press, 1980.

1991. Postscripts on survival and identity. In *Self and identity*, eds. Daniel Kolak and Raymond Martin, 285-289. New York: Macmillan, 1991.

Locke, John. 1690. *An essay concerning human understanding.* Oxford: Claredon Press.

Loeb, Louis. 1981. *From Descartes to Hume*. Ithaca: Cornell University Press.

Lombard, Lawrence. 1986. *Events: A metaphysical study*. London: Routledge.

Loux, Michael J. 2002. *Metaphysics: A contemporary introduction*. New York: Routledge.

Macdonald, Cynthia. 1989. *Mind-body identity theories*. London: Routledge.

2005. *Varieties of things: Foundations of contemporary metaphysics*. Oxford: Blackwell Publishing.

Macdonald, Cynthia and Graham Macdonald. 1986. Mental causes and the explanation of action. *Philosophical Quarterly* 36, 145-158.

1995. How to be psychologically relevant. In their *Philosophy of psychology: Debates in psychological explanation*, 140-157. Oxford: Blackwell, 1995.

2006. The metaphysics of mental causation. *Journal of Philosophy* 103, 539-576.

2007. Beyond program explanation. In *Common minds: Essays in honour of Philip Pettit*, eds. Geoffrey Brennan, Robert Goodin and Michael Smith, 355-372. Oxford: Oxford University Press, 2007.

Maxwell, Grover. 1962. The ontological status of theoretical entities. In *Minnesota Studies in the Philosophy of Science* 3, eds. Herbert Feigl, Michael Scriven, and Grover Maxwell, 3-15. Minneapolis: University of Minnesota Press, 1962.

McLaughlin, Brian. 2005. Supervenience. *Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/entries/supervenience/

Millikan, Ruth. 1990. Truth rules, hoverflies, and the Kripke-Wittgenstein Paradox. In *The philosophy of language*, ed. Aloysius Martinich, 545-561. New York: Oxford University Press, 2001.

1993. Explanation in biopsychology. In *Mental causation*, eds. John Heil and Alfred Mele, 211-232. Oxford: Oxford University Press, 1995.

1996. On swampkinds. *Mind and Language* 11, 103-117.

1999. Historical kinds and the "special sciences," *Philosophical Studies* 95, 45-65.

2008. Biosemantics. In *The Oxford handbook of philosophy of mind*, eds. Brian MacLaughlin, Ansgar Beckermann, and Sven Walter, 394-406. Oxford: Oxford University Press, 2008.

Musgrave, Alan. 2004. How Popper (might have) solved the problem of induction. In *Karl Popper: Critical appraisals*, eds. Graham Macdonald and Philip Catton, 16-27. London: Routledge, 2004.

Neander, Karen. 2004. Teleological theories of mental content. *Stanford Encyclopaedia of Philosophy*. http://plato.stanford.edu/entries/content-teleological/

2006. Content for cognitive science. In *Teleosemantics*, eds. Graham Macdonald and David Papineau, 167-194. Oxford: Oxford University Press, 2006.

Newman, Max 1928. Mr. Russell's 'causal theory of perception.' *Mind* 36, 137-148.

Papineau, David. 2002. *Thinking about consciousness.* Oxford: Oxford University Press.

2007. Must a physicalist be a microphysicalist? *Conference on Emergence*, Queen's University Belfast.

Pereboom, Derk. 1991. Why a scientific realist cannot be a functionalist. *Synthese* 88, 341-358.

2002a. On Baker's *Persons and bodies*. *Philosophy and Phenomenological Research* 64, 615-622. http://www.arts.cornell.edu/phil/homepages/pereboom/baker.html

2002b. Robust non-reductive materialism. *Journal of Philosophy* 99, 499-531.

Pereboom, Derk and Hilary Kornblith. 1991. The metaphysics of irreducibility. *Philosophical Studies* 63, 125-145.

Pettit, Philip. 1993. A definition of physicalism. *Analysis* 53, 213-223.

Plato. *Parmenides.* In *The collected dialogues of Plato*, ed. Edith Hamilton and Huntington Cairns. Princeton: Princeton University Press, 1961.

Poland, Jeffrey. 2003. Chomsky's challenge to physicalism. In *Chomsky and his critics*, eds. Louise Antony and Norbert Hornstein, 29-48. Oxford: Blackwell, 2003.

Psillos, Stathis. 1999. *Scientific realism: How science tracks truth.* London: Routledge.

Putnam, Hilary 1967a. The nature of mental states. In *Readings in philosophy of psychology* vol. 1, ed. Ned Block, 223-231. Cambridge: Harvard University Press, 1980.

1967b. Philosophy and our mental life. In *Readings in the philosophy of psychology* vol. 1, ed. Ned Block, 134-143. Cambridge: Harvard University Press, 1980.

1975. The meaning of 'meaning'. In *Language, mind, and knowledge*, ed. Keith Gunderson, 131-193. Minneapolis: University of Minnesota Press, 1975.

Quine, Willard. 1961. Reference and modality. In his *From a logical point of view*, 139-159. New York: Harper and Row Publishers, 1961.

1986. Reply to Morton White. In *The philosophy of W.V. Quine*, eds. Lewis Hahn and Paul Schilpp, 663-665. La Salle: Open Court, 1986.

Rappaport, William. 1994. Syntactic semantics. In *Thinking computers and virtual persons*, ed. Eric Dietrich, 225-274. San Diego: Academic Press, 1994.

Russell, Bertrand. 1927. *The analysis of matter.* London: Routledge.

1968. *The autobiography of Bertrand Russell* vol. 2. London: Allen and Unwin.

Sainsbury, Mark. 1995. *Paradoxes.* Cambridge: University Press.

Schwartz, Stephen. 1978. Putnam on artifacts. *Philosophical Review* 87, 566-574.

1980. Natural kinds and nominal kinds. *Mind* 89, 182-195.

1983. Reply to Kornblith and Nelson. *Southern Journal of Philosophy* 21, 475-479.

Searle, John. 1984. Intentionality and its place in nature. *Synthese* 61, 3-16.

1987. Indeterminacy, empiricism, and the first person. In *The philosophy of language*, ed. Aloysius Martinich, 484-497. Oxford: Oxford University Press, 2001.

1990. Is the brain's mind a computer program? In *Doing philosophy- An introduction through thought experiments*, eds. Theodore Schick and Lewis Vaugh, 114-121. Mountainview: Mayfield Publishing Company, 1990.

1993. The critique of cognitive reason. In *Readings in philosophy and cognitive Science*, ed. Alvin Goldman, 833-848. Cambridge: MIT Press, 1993.

1995. *The construction of social reality*. London: Penguin.

Shoemaker, Sydney. 1975. Functionalism and qualia. In *Readings in philosophy of psychology* vol. 1, ed. Ned Block, 251-267. Cambridge: Harvard University Press, 1980.

Sklar, Lawrence. 1982. Saving the noumena. In *The philosophy of science*, ed. David Papineau, 61-81. Oxford: Oxford University Press, 2001.

Smart, John. 1962. Sensations and brain processes. In *The philosophy of mind*, ed. Vere Chappell, 160-172. Upper Saddle River: Prentice-Hall, 1981.

Spelke, Elizabeth and Katherine Kinzler. 2007. Core knowledge. *Developmental Science* 10, 89-96.

Stairs, Allen. 1991. Quantum mechanics and the self. In *Self and identity*, eds. Daniel Kolak and Raymond Martin. New York: Macmillan, 1991.

Stich, Stephen. 1984. Could man be an irrational animal? Some notes on the epistemology of rationality. In *Naturalizing epistemology*, ed. Hilary Kornblith, 337-358. Cambridge: MIT Press, 1994.

1994. A pragmatic approach to cognitive evaluation. In *Naturalizing epistemology*, ed. Hilary Kornblith, 393-426. Cambridge: MIT Press, 1994.

Tarski, Alfred. 1944. The semantic conception of truth and the foundations of semantics. In *The philosophy of language*, ed. Aloysius Martinich, 61-91. Oxford: Oxford University Press, 2001.

Alan, Turing. 1936. On computable numbers, with an application to the *Entscheidungsproblem*. In *The essential Turing*, ed. Jack Copeland, 58-90. Oxford: Claredon Press, 2004.

1948. Intelligent machinery. In *The essential Turing*, ed. Jack Copeland, 410-431. Oxford: Claredon Press, 2004.

1950. Computing machinery and intelligence. In *The essential Turing*, ed. Jack Copeland, 441-464. Oxford: Claredon Press, 2004.

Tye, Michael. 1997. The representational character of pain. In *The nature of consciousness: Philosophical debates*, eds. Ned Block, Guven Guzeldere and Owen Flanagan, 329-340. Cambridge: MIT Press, 1997.

van Fraassen, Bas. 1980. *The scientific image.* Oxford: Claredon Press.

Williams, Bernard. 1970. The self and the future. In *Self and identity*, ed. Daniel Kolak and Raymond Martin. Upper Saddle River: Prentice-Hall, 1991.

Williams, D.C. 1953. The elements of being. In *Metaphysics: Contemporary readings*, ed. Michael Loux, 57-64. London: Routledge, 2001.

Wilson, Jessica. 2006. On characterizing the physical. *Philosophical Studies* 131, 61-99.

Worrall, John. 1989. Structural realism: The best of both worlds? In *The philosophy of science*, ed. David Papineau, 139-165. Oxford: Oxford University Press, 2001.

Yablo, Stephen. 1990. The real distinction between mind and body. *Canadian Journal of Philosophy* 16, 149-201.

1992. Mental causation. *The Philosophical Review* 101, 245-280.