INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

A

TOWARD A METAPHYSICS OF MENTAL CAUSATION

by

ROBERT C. BUCKLEY

A dissertation submitted to the Graduate Faculty in Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2001

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

01/23/01
Date

Chair of Examining Committee

01/23/01
Date

Executive Officer

Supervisory Committee:

Prof. Alan Hausman
Prof. John Greenwood
Prof. Charles Landesman
Prof. Michael Levin
Prof. Richard Mendelsohn

THE CITY UNIVERSITY OF NEW YORK

THE CITY UNIVERSITY OF NEW YORK

Abstract

TOWARD A METAPHYSICS OF MENTAL CAUSATION

by

Robert C. Buckley

Adviser: Professor Charles Landesman

In this dissertation I defend antireductive physicalism against the charge that it cannot provide a workable account of mental causation. I consider two forms of antireductive physicalism: Donald Davidson's anomalous monism (AM) and functionalism. The criticism that has been raised against these views is that, to the extent that they make mental properties irreducible to physical properties, they render mental properties causally impotent with respect to the physical world.

My response to this attack is two-fold. In the first place I distinguish between classificatory mental properties and non-classificatory mental properties. Both AM and functionalism claim that general classificatory

properties like being a desire, being a belief that p, being a pain, etc. cannot be reduced to the properties of physical theory. But how things are classified should not be expected to affect the causal efficacy of those things. Indeed, the classifications of the functionalist are based on the *prior* causal capacities of the physical events and properties that count as realizations of mentality. Hence, irreducibly mental properties of a classificatory sort do not, but also *should not*, be expected to have any causal efficacy.

Non-classificatory mental properties however, like the painfulness of a burn or the itchiness of a mosquito bite ought to be expected to play a causal role if folk psychology is correct. But if these non-classificatory mental properties are identical with instances or "tropes" of physical properties, then there should be no special worry about the causal efficacy of the mental for the antireductive physicalist. Since such a property identification move is open to the antireductive physicalist, I argue that he has no special problem accounting for mental causation despite the protests of Jaegwon Kim, Ernest Sosa, Ted Honderich and others.

Acknowledgments

I would like to thank those individuals without whom successfully completing this dissertation would have been impossible. First of all, I owe a special debt of gratitude to my adviser Jim Landesman whose generosity, wisdom, and encouragement made this thesis a reality. I feel very fortunate to have had the opportunity to work closely with a philosopher of such exceptional caliber.

Next, I must thank Maureen Eckert for her patient and careful reading of many early drafts of this thesis. Her encouragement and critical feedback were instrumental in the formation of the final draft.

I would also like to thank Robert Talisse for his insightful comments on chapter six and his insistence on following the *logos* at all times.

Lastly, I must thank my mother Anne Maviglia and grandmother Felicia Maviglia who gave me the courage to become a philosopher. This thesis is dedicated to them.

TABLE OF CONTENTS VII

Chapter I: Setting Up the Problem of Mental Causation

1. Cartesian Substance Dualism

Since at least the publication of Descartes' *Meditations on First Philosophy*[1] philosophers have tried to provide an account of the mind-body relation that both preserves the distinctness of the mental from the physical and, at the same time, explains the apparently strong causal tie that binds together minds and bodies. Minds seem

[1]Rene Descartes, *Meditations on First Philosophy*, in *The Philosophical Writings of Descartes, vol. II*, edited by John Cottingham, Robert Stoothoff, and Dugald Murdoch (Cambridge University Press: Cambridge, 1984).

fundamentally different from bodies in many respects but they also seem to be tightly linked with them. As Descartes himself described it, there appears to be an intimate relationship between the two:

> Nature...teaches me, by these sensations of pain, hunger, thirst and so on, that I am not merely present in my body as a sailor is present in a ship, but that I am closely joined and, as it were, intermingled with it, so that I and the body form a unit.[2]

Descartes thought that minds were quite different from bodies insofar as they are capable of thought and consciousness. Minds think, feel, and experience consciousness generally, but bodies do not. But regardless of how special or unique minds seem to be when compared to bodies, they are nevertheless tightly linked with them. When I burn my hand, I feel a pain. When my body needs nourishment I feel hungry. When I desire to move my arm, more often than not, my arm moves. It seems obvious that

---

much of what happens to our bodies has an effect on our minds and that our beliefs and desires also have an impact on our bodies.

Unlike a sailor in a ship, each of us seems to be intimately connected to many of the goings on of our bodies. We are not mere spectators on the realm of physical existence--we seem to be linked to it through various kinds of intimate connections. We may "steer" our bodies through the world in a manner analogous to that in which a pilot steers a ship, but unlike a pilot, we have a more direct connection to our "ships." When our bodies are damaged, we feel pain. Hence the connection between mind and body is quite close, as Descartes suggests in the above quotation. Is there is way to explain what is distinctive about the mind, and at the same time account for the fact that minds seem to affect bodies?

According to Descartes, material bodies are essentially distinct from minds. The essence of body is extension, while the essence of mind is the activity of thinking. Descartes infers from this a metaphysical distinction between the two: Material substances take up space and can

---

[2] *Ibid.*, p.56.

be divided, but mental substances, by their very nature, are indivisible and do not seem to take up space.[3] In the Sixth Meditation he writes:

> …[T]here is as great a difference between the mind and the body, inasmuch as the body is by its very nature always divisible, while the mind is utterly indivisible. For when I consider the mind, or myself in so far as I am merely a thinking thing, I am unable to distinguish any parts within myself. Although the whole mind seems to be united with the whole body, I recognize that if a foot, or arm, or any other part of the body is cut off, nothing has thereby been taken away from the mind.[4]

As Margaret Wilson formulates it, Descartes' argument for mind body dualism goes as follows.[5] Descartes notes that he has a clear and distinct idea of his body as distinct from his mind. His mind seems indivisible, whereas

---

[3] Though they may exist at a point, in the pineal gland, for Descartes.
[4] Ibid., p.58

his body is certainly divisible. And since the existence of God ensures that any such clear and distinct ideas must correspond to reality, Descartes concludes that this distinction is a real distinction, i.e. that corresponding to his distinct concepts of mind and body there must be a metaphysical distinction between mental and physical entities. That it is to say, minds and bodies are not only conceptually distinct, but they are distinct entities. This is the essence of mind/body dualism.

The validity of this argument is doubtful. For it is far from clear why, assuming Descartes' line of reasoning, he is entitled to infer that there must be a metaphysical distinction where there is a conceptual one. For example, one's concept of the morning star may be distinct from one's concept of the evening star despite the fact that these concepts hold of the very same object. Why the same cannot be said of the mental-physical contrast here is unclear. Nevertheless, Descartes did accept the notion that minds and bodies are distinct entities.

According to this view then, minds are substances, distinct in kind from all material objects. They are

---

[5] In her *Descartes*, (Routledge: New York, 1978), chapter VI.

indivisible and non-spatial. But as Gassendi pointed out in Descartes' own time, this leaves us with a perplexing account of mental causation. He challenges Descartes on this score as follows:

> The...problem...arises concerning the spirits which you must transmit in order to communicate feelings or instructions, or to initiate movement...you must explain to us how this 'directing' of movement can occur without some effort--and therefore motion—on your part. How can there be effort directed against anything, or motion set up in it, unless there is mutual contact between what moves and is moved? And how can there be contact without a body?[6]

In other words, if Descartes is right about the mind being incorporeal it is difficult to understand how he can account for mental causation. For it is mysterious on this view how my mind, for example, lacking spatial properties,

---

[6] Cottingham, J. et al (eds.): *The Philosophical Writings of Descartes, vol. 2* (Cambridge University Press: Cambridge, 1984), p.237

might somehow reach into the physical world and stimulate a neural firing that gets my body to move. Does it somehow rub up against neurons in my brain, triggering a motor response from the rest of my body despite the fact that this mind and its thoughts and desires have no mass, velocity, or magnitude?

In being so distinct from bodies, mental substances enjoy a very mysterious relation to bodies and the physical realm generally. The tight link between mind and body that Descartes acknowledged is in conflict with his dualistic theory of mind.

Descartes' own notorious attempt to answer to these questions was that the pineal gland somehow serves as the locus for mind-body interaction. But of course shifting the site of such interaction from the whole body or brain to one particular, albeit very small part of the body does very little to illuminate these matters. If Descartes and much of common sense is right, and there is genuine causal commerce between one's mind and one's body, between the mental and the physical generally, then Descartes' mind-body dualism simply won't do. We must look elsewhere for an account that will explain these causal facts.

Descartes' problem, at least according to this line of criticism, is that he posited a metaphysical gulf between the mental and physical domains. Today, most philosophers maintain some form of physicalism, according to which there are only physical causes and effects of events in the universe. According to this view, the world is thoroughly physical at root, and as such, it is causally closed. This means that mental substances may not creep into the causal nexus or somehow be squirted out of it, if this ontological claim is correct. Physicalism has the virtue of avoiding a sharp cleavage between the mental and the physical and so the problem of causal interaction would seem less of a worry for physicalist theories of the mind.

The most straightforward form of physicalism to be found in the philosophy of mind is perhaps the mind/brain identity thesis according to which, mental events are simply brain processes. On this view, even if mental terms do not have the same meaning as neurophysiological terms, they are at least co-extensional with them in the same way that the terms 'water' and 'H2O' are co-extensional even though they have different meanings. Someone may sincerely believe that water, and not H20, is a potable liquid that can be found in

rivers, lakes, and streams, for example, despite the fact that the two terms, 'water' and 'H20' have the same referent. This may be the case because the term 'water' picks out the same stuff as the term 'H20', but in a different way. One needs to learn that water is H20 to see that the two terms pick out the same substance.

Nevertheless, water is nothing "over and above" H2O it is sometimes said, and in the same way, mental states, like pain, may be nothing "over and above" various goings on in the central nervous system. The precise nature of such mental-physical identities, it is thought, should be a matter of empirical research in neurophysiology. But when we refer to psychological events like seeing a blurry after image, or feeling a sharp pain, for example, we are referring to some brain process, even if we are unaware of this reference. Here is how J.J.C Smart described the thesis in his "Sensations and Brain Processes:"

> It is not the thesis that, for example, "after-image" or "ache" means the same as "brain process of sort X" (where X is replaced by a description of a certain sort of brain process). It is that,

> in so far as "after-image" or "ache" is a report of a process, it is a report of a process that *happens to be* a brain process. It follows that the thesis does not claim that sensation statements can be *translated* into statements about brain processes. Nor does it claim that the logic of a sensation statement is the same as that of a brain-process statement. All it claims is that in so far as a sensation report is a report of something, that something is in fact a brain process. Sensations are nothing over and above brain processes.

According to this view, it might seem odd to describe a brain process as dull or stabbing. However, this should not pose any obstacle to the idea that such neural phenomena happen to be identical with mental states like aches and pains, which may properly be described as dull or stabbing. For the meanings of mental terms can be distinct from the meanings of brain process terms even if we discover that such terms refer to the very same entities. Smart argued

that we should expect type-type identities between the mental and the physical whereby types of mental phenomena, like pains, are shown to be identical with some given type of physical phenomenon, like C-Fiber firings.

As a theory of the mind-body relation, the identity thesis has at least one obvious strength. If one opposes Cartesian dualism on the ground that it cannot account for causal commerce between the mind and the body because of the metaphysical divide that separates them, then the identity theory will appear an attractive alternative. If a mental event is simply identical with some brain event, the problem of causal interaction would seem to dissolve: it is no worse a problem than that involved in understanding any other species of physical, causal interaction. Understood this way, mental causation would seem relatively unproblematic.[7]

2. The Token Identity Thesis

The mind-brain identity theory did not enjoy long-lived success however. Part of the problem lay in the fact that it required empirically implausible identities. For

---

[7] I say "relatively unproblematic" because the nature of causation has been a vexing philosophical problem since Hume's *A Treatise of Human Nature*.

example, if the identity theorist is right, then pain must be identical with some particular class of neurophysiological phenomena. Hilary Putnam argued that such identities will probably turn out to be false as a matter of empirical fact when we consider the "multiple realizability" of the mental.[8] Considering the hypothesis that mental states simply are brain states (the type identity thesis) he writes: [9]

> Taking the brain-state hypothesis in this way then...consider what the brain state theorist has to do to make good his claims. He has to specify a physical-chemical state such that *any* organism (not just a mammal) is in pain if and only if (a) it possesses a brain of a suitable physical-chemical structure; and (b) its brain is in that physical-chemical state. This means that the physical-chemical state in question must be a possible state of a mammalian brain, a reptilian

---

[8] See his "The Nature of Mental States," reprinted in *The Nature of Mind*, edited by David Rosenthal (Oxford University Press: New York, 1992) pp.197-203.

[9] Ibid, pp.197-198.

> brain, a mollusc's brain (octopuses are molusca, and certainly feel pain), etc. At the same time, it must *not* be a possible (physically possible) state of the brain of any physically possible creature that cannot feel pain. Even if such a state can be found, it must be nomologically certain that it will also be a state of the brain of any extraterrestrial life that may be found capable of feeling pain before we can even entertain the proposition that it may *be* pain.

The idea here is that a mental state, like pain, might be realized by one type of physical processes in humans, some other set in mollusks, and perhaps an even more exotic set in extraterrestrial beings, if any should exist; and so on, possibly indefinitely.

If pain is like this, then what it is to be a state of pain will be multiply realizable in the way that being an adding machine is multiply realizable. A computer made of plastic, silicon chips, and transistors might serve as an

adding machine; but so might an abacus made of glass beads and wooden rods. Yet each of them is an adding machine despite these differences in physical make-up.

If pain is similarly multiply realized, then the identity thesis must be wrong. For the identity thesis places a strict requirement on any possible realization of pain: it must be a brain state of a particular physical-chemical structure, as Putnam points out. But this is too strong because we cannot know *a priori* whether every creature that would seem to meet a functional definition of pain would also share our neurochemistry. This *may* be the case, but it seems very unlikely. Part of Putnam's claim here is that we can decide whether something experiences pain independently of knowing anything about its physical make-up. Hence the question of whether it does experience pain is not determined by questions concerning its physical make-up.

If pain is defined functionally, then we might decide whether a creature experienced pain not by inspecting its neurophysiology, but by considering its overall "functional

economy." [10] The idea here is that pain will be defined as some state in a creature that plays a role in that creature's interactions with the environment such that:[11]

> …[T]he functional state we have in mind is characterized, at least partially, by the fact that the sense organs in question are organs whose function is to detect damage to the body, or dangerous extremes in temperature, pressure, etc., and by the fact that the "inputs" themselves, whatever their physical realization, represent a condition that the organism assigns a high disvalue to…

Or consider David Lewis' functional account of pain and mental phenomena generally:[12]

> Our view, is that the concept of pain, or indeed of any other experience or mental state, is the

---

[10] Putnam, Hillary, "The Nature of Mental States," p. 202 in *The Nature of Mind*, David Rosenthal (ed.).

[11] *Ibid.*

[12] Lewis, David. "Mad Pain and Martian Pain," in *The Nature*

> concept of a state that occupies a certain causal role, a state with typical causes and effects. It is the concept of a state apt for being caused in certain ways by stimuli plus other mental states and apt for combining with certain other mental states to jointly cause behavior...If the concept of pain is the concept of a state that occupies a certain causal role, then whatever does occupy that causal role is pain. If the state of having neurons hooked up in a certain way and firing in a certain pattern is the state properly apt for causing and being caused, as we materialists think, then that neural state is pain.

In short, if we define pain in our own case functionally, then we might consider it to be that state in humans that does all of the sorts of things Putnam and Lewis list above. But if creatures with very different physical make-ups have functionally similar states, they should also count as pain-feeling organisms.[13] In their case, something

---

of Mind, David Rosenthal p. 230.

[13] Please note there is nothing in Putnam's argument that

very different from the neural structure that normally occupies the functional role of human pain may occupy the functional role of that state.[14] In an analogous way, something might count as a chair whether it is made of wood, plastic, or bronze. Notice then, the question of whether creatures experience pain is divorced from questions concerning their physiology in just the way that the question of whether a given object is a chair is divorced from the question of what it is made of.

This suggests that psychology is at least explanatorily autonomous from physiology in that explanations concerning mental states like pain may not make any significant reference to the physiological facts of pain-feeling systems. But one might insist that despite this "functional" aspect of the mental and its corresponding cleavage between psychological and neurophysiological descriptions, there must be some straightforward dependence of the mental on the

---

limits the notion of pain to biological creatures. Hence 'organism' should be understood as referring to biological, as well as, non-biological systems here.

[14] As Antony and Levine point out in their "Reduction with Autonomy" (1997) it could also turn out that there might be creatures who experience C-fiber firings, but for whom such events are causally responsible not for feeling pain sensations, but for "feeling groovy."

physical in order to account for causal interactions between the mental and physical domains.

One can then turn such a functionalist characterization of the mental into a version of the "token identity thesis," according to which every mental state token must be identical with some particular physical token. One can do this and deny that there must be identities between types of mental states, like pain, and types of physical states, like C-Fiber firings, thereby accommodating the multiple realization claim.

Let me say a few words about the token identity thesis before proceeding. According to this view, every mental event must be identical with some particular, dateable, physical event, though there is no reason to expect there to be any type of physical event common to all instances of particular classes of mental events, like experiencing a pain. An analogous token identity obtains between being a colored object and having a shape.[15] Every colored object,

---

[15] The example that follows is based on Jaegwon Kim's discussion in "Psychophysical Laws," (1987). It must be noted that this example does not provide a perfect analogy between with the mental/physical case. This is so because the mental/physical case is one of supervenience, whereas the color/shape case is not. The point though is that it is possible to have a token identity without a type identity,

we might note, must have a shape. But we know that not every such object has the same shape.

So, while every colored object must instantiate a "shape token" it does not follow that there will be matches between types of color and types of shape, e.g. all red things being round.

Analogously, every psychological state will be identical with some physical state. But even if every occurrence of a psychological state like pain must be the occurrence of some physical event, it does not follow that it must be *the same type* of physical event in all such occurrences. C-fiber firings may count as pains in humans, while something very different physically, say M-fiber firings, may count as pains in Martians.

Moreover, and perhaps more importantly, such diversity does not require a metaphysical gap between the mental and the physical. For even if M-fibers are different sorts of things from C-fibers, they are still physical. There is no need to suppose that pains must somehow be non-physical simply because their various realizations share little in

---

and this is brought out by the shape/color analogy. More will be said of supervenience in later chapters.

common physically. We do not think chairs are non physical simply because they have myriad physical realizations. Without such a gap between the mental and the physical, mind-body causal interaction would seem unproblematic.

Donald Davidson has offered a physicalist alternative to the type identity view. His position, Anomalous Monism, is a version of token physicalism according to which, every mental event is identical with some physical event, though he denies that we should expect any type identities between mental events and physical events. So, my desiring a dry martini must be a physical event, on Davidson's view, though there is no reason to think that every time I, or anyone else desires a dry martini, some type identical physical event is occurring within us. One neural configuration might realize the desire on one occasion, and some other, very different neurological configuration might do so on another occasion. You and I might both desire a dry martini even if we are in quite different neurophysiological states.

Like the fan of multiple realizability, the anomalous monist claims that all psychological phenomena must be physically realized in order to account for the causal commerce that obtains between the mental and the physical.

But there is no insistence that mental properties like 'being a desire' must be identical with any specifiable physical property or determinate set of physical properties. On the contrary, the anomalous monist endorses the notion that psychological kinds cannot reduce smoothly to physical kinds and hence, that certain mental properties cannot be reduced to physical properties.[16] Although they hold this view for different reasons,[17] the Davidsonian and the functionalist can be seen as advocating a roughly similar account of the mind-body relation insofar as both endorse some form of the token identity thesis. Unlike Descartes, they both hold that the mental is physical. But they deny that there will be type identities between mental events and physical events. Hence they deny the reducibility of mental properties to physical properties.

3. Stating the Problem of Mental Causation

---

[16] More will be said on just what this claim amounts to in chapters III and V.

[17] In chapter V I will discuss in greater depth Davidson's reasons for making this claim.

In what follows I will be addressing the concern raised by many philosophers[18] that such an account of the mind body relation, like Descartes', founders on the problem of mental causation. According to this line of criticism, anomalous monism is really epiphenomenalistic since it makes the distinctly psychological properties of events, like being a desire, irrelevant from a causal point of view. The idea, in a nutshell, is that physical descriptions of events capture their lawful regularities and their causal interconnections, whereas mental descriptions of the same events do not capture these regularities. Hence it is only as physical events that events like desiring a dry martini figure into causal laws, and hence, enter into causal relations, according to these critics.

But if being psychological events makes no difference to the causal relations into which desires and other intentional states enter, it would seem as if these mental

---

[18] C.f. e.g., F. Dretske, "Reasons and Causes," *Philosophical Perspectives* 3 (1989), pp. 1-15; T. Honderich, "The Argument for Anomalous Monism," *Analysis*, 42 (1982), pp.59-64; E. Sosa, Mind-Body Interaction and Supervenient Causation," in P. French *et. al* (eds.) *Midwest Studies in Philosophy, IX* (University of Minnesota Press, 1989), pp. 271-81; and J. Kim, "Epiphenomenal and Supervenient Causation" in his *Supervenience and Mind* (Cambridge University Press: Cambridge, 1993), pp. 92-108.

properties are epiphenomenal. The basic idea is that even if mental events are physical events, their being mental plays no causal role. To see what is at issue, compare the following two causal statements:

1. This 50kg. weight crushed John's toe.
2. This 50kg. weight, *which was bought as a birthday present for Samantha*, crushed John's toe.

Clearly, being a birthday present for Samantha is a property that plays no causal role with respect to the effect in question here. The argument against Davidson's Anomalous Monism suggests that the view fails because it makes mental properties inefficacious in an analogous way. For example, if the event that is my desiring a beer is also a physical event, then according to the critics of anomalous monism, it is unclear why we should think that being a desire figures at all into causal explanations that refer to this event.

After all, it seems that a neurophysiological account could tell us all about the causal linkages that result in the movement of Jones's body towards a beer without mentioning anything about his desire. Even if his desire

happens to be token identical with some state/event of his brain then the fact that such a state/event is mental would seem to do no causal work. Being a birthday present of Samantha is token identical with being the 50 kg. weight that crushed John's toe even though it is of no causal relevance to that effect.

But if common sense is right and our bodies often move, in part, because of what we think and feel, this result is intolerable, for it suggests that such mental events do not cause anything, as mental events. At least this is what the critics have claimed against Davidson. Since Davidson is famous for defending the thesis that reasons can be causes of behavior, this consequence would seem particularly problematic for him.[19]

As non-reductive physicalists, both the Davidsonian and the functionalist seem to face a nasty problem. On the one hand, they are *physicalists* and as such, maintain some version of the Causal Closure Principle, according to which there are only physical causes and effects of events in the world. They both preclude the possibility of Cartesian

---

[19] See his "Actions, Reasons, and Causes," (1963) reprinted in his *Essays on Actions and Events* (Oxford: New York, 1980) pp.3-21.

interactionism, vitalism, or any ontological mixture of physical and non-physical entities[20] interacting causally with one another.

On the other hand, as *non-reductive* physicalists they also deny that there is an identity between mental properties and physical properties. The functionalist does so by citing the multiple realizability of the mental: being a pain, or any other type of psychological state is irreducible to any given physical property. And for Davidson, being an intentional state is conceptually irreducible to any given physical property because he thinks that attributions of intentional states rely essentially on normative constraints that cannot be accommodated within the conceptual apparatus of physics.[21]

The problem then becomes one of explaining the causal efficacy of mental properties within a physical world that

---

[20] By "entities" I mean to include properties as well as states, events, and objects.

[21] Davidson restricts his treatment of mental properties to intentional states like belief and desire. It is *these* states/events, not phenomenal states/events that are conceptually irreducible to the physical. Again, spelling out Davidson's reasons for holding the so-called "anomalousness" of the mental is part of the aim of chapter four.

is causally closed. To make clear just what is at issue, it will be helpful to describe the problem in terms of four claims made by the non-reductive physicalist, which seem, *prima facie*, inconsistent:

1. **The Causal Closure Principle:** The physical world is causally closed, i.e. there can be no bleeding in or out of the causal nexus of physical interactions. This means that non-physical entities do not have causal commerce with the physical.

2. **The Psychophysical Interaction Thesis:** Physical states sometimes cause, and are caused by, mental states.

3. **The Anti-Reductive Thesis:** While token mental states may be identical with various physical tokens, mental properties, like being a pain, are not reducible to physical properties, like being a C-fiber Firing.

4. **The Mental Causation Thesis:** When mental events bring about physical effects, they do so *in virtue of (at least some of) their mental properties.*

The problem of mental causation for the non-reductive physicalist then comes to this: to explain how the irreducibly mental properties of mental states/events can be causally efficacious in a causally closed, physical world. If they are efficacious, though not identical with any physical property, proposition #1 is false and causal closure is violated. Non-physical properties would seem as problematic as Cartesian substances for any adherent of the Causal Closure principle. But if such properties are identical with some physical property or other, we gain causal interaction at the price of giving up #3, in which case non-reductionism is sacrificed. The challenge of non-reductive physicalism is to show how, despite appearances to the contrary, these four statements might be jointly true. On the surface it seems a Herculean task.

4. Two Types of Property Dualism

Before proceeding I think it is important to note exactly what sort of irreducibility claim both Davidson and the functionalist are making, for this will figure importantly in assessing the validity of any proposed

solution to the problem of mental causation outlined above. They are, it seems, concerned primarily with properties like *being a desire, being a pain, being a belief that p,* etc.—properties that define primarily *classes of mental phenomena*.[22] Human pains may consist solely of C-fiber firings, but dogs might lack entirely C-fibers. Hence, if we accept the functionalist characterization of pain we must accept that dogs experience pain as long as there is *some state* in them which plays the functional role characteristic of pain. For example, if D-fibers are responsible for getting dogs to become aware of tissue damage or engage in avoidance behavior when such damage occurs, or to cry, whimper, etc. then we must count D-fiber firings as pains.

So, the functionalist gets us to ignore the specifics of human pain because the *concept* of pain, according to this line of thought, is a functional concept. Anything that plays a role in the overall economy of a system such that it is isomorphic to that played by C-fibers in humans counts as pain, for the functionalist. But notice it is the *concept of*

---

[22] It is important to note again that Davidson's AM is concerned only with intentional properties—not phenomenal properties.

pain that cannot be reduced physically.[23] As a physicalist, the functionalist is not arguing that pains must exist in some non-physical way, perhaps lacking spatial properties, as a Cartesian would insist.

So a mental property, like *being a pain*, would count as irreducible on the functionalist view because there is nothing that binds together all instances of pain as such that would form a natural grouping at the physical level. Spelling out this notion in more detail is the subject of chapters II and III in which I examine the notion of reduction and the functionalist claim that mental properties are not reducible to physical properties. But we can get an intuitive grasp of this idea here with the following example.

Consider, for instance, what it is for something to count as water. To be an instance of water is just to be an instance of H20. Hence all, and only water, is H20. By contrast, what it is for something to be a delicious drink might not be so reducible because there may be no natural grouping at the physical level that captures, or maps on to, the property *being a delicious drink*. We may not be able to

---

[23] Spelling out this notion of reducibility in greater

get a physical description of such a concept because there may be no determinate set of molecular properties that exhausts all possible instances of being a delicious drink.

Being a mental state of some kind, according to functionalism, is more like being a delicious drink than being water--the patterns distinctive of such states just do not admit of a clean reduction to any lower level, such as that studied by physics or neurobiology. The physical level might be able to tell us all about the neurological mechanisms that inform human pain without capturing the patterns distinctive of pain in general, whether it is human pain or some more exotic variety.

One way to understand this idea of "capturing" is to consider it in terms of a property or set of properties that can be articulated in some non-arbitrary way at the physical level. If all pains were C-fiber firings, this would be accomplished. For in that case, the relation between being a pain and being a C-fiber firing would be like the relation between being water and being H20. It would be a matter of having uncovered all realizations of pain in physical or neurobiological terms.

---

detail is the subject of chapter III.

But if our concept of pain is functional, then there may be no non-arbitrary limit on what physical entities are to count as instances of pain. So even if a long list of all known pain realizations were compiled, *what makes them pains* would be lost in such a list. Why these things all fall on *this* list, and not some other, may not be explicable from the physical point of view because of wide differences in the physical natures of these realizations.[24]

Similarly, what new members of the list to include would be an issue not resolvable from within the purview of physical theory. If pains are defined functionally then we must decide whether something is a pain in isolation from its physical properties. Hence properties like *being a pain* do not find a smooth reduction into physical theory if they are primarily functional concepts. Again, spelling out this

---

[24] See Ned Block's "Introduction: What is Functionalism?" for a forceful presentation of this point in Ned Block, ed. *Readings in the Philosophy of Psychology*, vol. 1 (Harvard University Press: Cambridge, 1980). For a more recent defense of this claim, see also Jerry Fodor's "Special Sciences: Still Autonomous After All These Years," in *Philosophical Perspectives, Vol. 11: Mind, Causation, and World*, (Blackwell: Boston, 1997). Also see Jaegwon Kim's "Multiple Realization and the Metaphysics of Reduction," in *Philosophy and Phenomenological Research* 52, pp. 1-26 for an argument against this claim.

claim in detail will be taken in up in the first two chapters to follow.

Davidson makes a similar claim about the irreducibility of the mental. If being a desire is not something that physical theory can tell us about—-if what makes something a desire, or the desire for some specific x still cannot be determined by a physical analysis, then properties like *being a desire* would be irreducible because of this conceptual asymmetry. This is just what Davidson claims is the case. In chapter V we will take up the reasons why he holds this to be true and I will argue that the reasons for which Davidson believes the mental to be irreducible to the physical also pose a problem for his account of mental causation.

Each of the above-described views of the mental/physical contrast amounts to a form of property dualism. For in each case, mental properties are understood to be independent in some sense, of physical properties. What it is to count as a pain, for instance, is independent of the physical facts, even if there can be no pain states in the absence of physical states.

But this property dualism must be contrasted with another, more metaphysically exotic type of property dualism. There is a set of arguments in philosophy of mind that attempt to show, in one way or another, that the qualitative aspects of experience just cannot be explained in physical terms. Thomas Nagel[25] and Frank Jackson[26] have both argued that mental properties are essentially subjective and hence any attempt to identify them in a type *or token fashion* with physical properties, is doomed to failure. Jackson and Nagel suggest that we simply cannot make sense of the notion that subjective feels, like those involved in the experience of seeing red or of navigating through the air by sonar, could be identical with any physical goings on in a creature's nervous system. Hence they pose a challenge to the very notion of physicalism which undergirds the debate between the non-reductive and reductive physicalists, i.e. between the token physicalist of either the functionalist or anomalous monist varieties,

---

[25] Nagel, Thomas, "What Is It Like to Be a Bat?" in David Rosenthal (ed.), *The Nature of Mind* (Oxford: New York, 1991), pp. 422-429.

[26] Jackson, Frank, "What Mary Didn't Know," in David Rosenthal (ed.), *The Nature of Mind*, (Oxford: New York, 1991), pp. 392-395.

and the type physicalist. For Jackson and Nagel, the very idea that mental properties might be physical, in any sense, is problematic.[27]

Though I will argue in chapter VI that the Nagel-Jackson line is importantly confused, in the first four chapters of this dissertation, I will simply assume a broadly physicalist framework. This is so because the question of mental causation, as I am considering it here, is really a dispute between different varieties of physicalism. It is an "in-house" debate over the viability of a certain version of the physicalist view.

The controversy arises upon consideration of the following facts. On the one hand, it seems we need to show how it is that mental properties play a real causal role in our explanations of behavior. For instance, we need to show how it is that Mary's *wanting to raise her hand* was causally relevant to her arm rising. If the causal story can be told without any reference to such mental phenomena, it would seem that the mental does no causal work in such cases.

The reason this is thought to be a particularly

---

[27] Though, as we will see in chapter II, Jackson allows for the possibility that mental states could be identical with

difficult problem is related to a second widely accepted claim. The world of causes and effects is physical through and through. This is one way of expressing the Causal Closure Principle alluded to above. According to this claim, there can be no creeping into or out of the causal nexus by entities that are not physical.

Hence, if mental properties are distinct from physical phenomena, as antireductionists would have us believe, it is hard to see how they could play any causal role in the world at all. If, as Lousie Antony and Joseph Levine proclaim[28] "we want the mental facts to be both real and different from the physical facts" but, at the same time, we want to maintain the causal closure of the physical and the reality of mental causation, then we have a serious philosophical problem to address. To the extent that they are irreducible, mental properties seem cut off from the pushing and shoving that goes on at the physical level.

---

physical states after all.

[28] Antony, Louise and Joseph Levine. "Reduction with Autonomy," in *Philosophical Perspectives, Vol. 11: Mind,*

Causation, and World (Blackwell: Boston, 1997) pp.81-105.

Chapter II: Recent Attempts to Solve the Problem of Mental Causation

The problem of mental causation as we have been discussing it seems intractable. If one accepts that only physical properties may have causal efficacy within the physical world, and that mental properties are distinct from physical properties, it would seem outright irrational to insist further that mental properties play a causal role in the physical world. Nevertheless, common sense seems to demand this. For I often do what I do because of what I think, feel, desire, etc. I scratch an itch because it feels a certain way, I raise my hand because I want to make a bid, I say what I think because I want to express my thoughts, and so on.

In this chapter I survey some recent attempts to provide a philosophical basis for this common sense intuition. The first such attempt is an intriguing suggestion put forth by Frank Jackson and Philip Pettit. Despite powerful considerations which suggest that mental properties have no causal efficacy, Jackson and Pettit claim that such properties may have causal *relevance* because they serve to "program for" their effects. In a nutshell, they argue that the presence of a mental property can ensure that a certain range of effects will occur, and as such, mental properties are relevant to their effects. This is true even if mental properties are not themselves efficacious.

Ernest LePore and Bary Loewer attempt a similar resolution of the mental causation problem. Their approach also consists in trying to secure the causal *relevance* of mental properties. But they try to show that counterfactual statements such as "If Sheila had not wanted to drink some water she would not have extended her hand" can provide the basis for the sense of relevance necessary for folk psychology. LePore and Loewer think that such statements show that if a certain mental property did not exist at a given time, then the precise effect that arose would not

have arisen. Hence, mental properties can be relevant even if they do no causing themselves.

After arguing that neither of these solutions provides a sense of causal relevance robust enough to save folk psychology from epiphenomenalist worries, I will turn to a recent strategy in the literature to explain away the need for a metaphysical solution to the problem of mental causation. Lynn Rudder baker and Tyler Burge have both argued that the explanatory success of folk psychology should be used a basis for rejecting the metaphysical assumptions that give rise to the problem of accounting for the causal relevance of mental properties.[29] Since we have more confidence in the reality of mental causation than we do in the various metaphysical assumptions that are thought to give rise to the problem, we have an argument for "dissolving" the problem and the assumptions which give rise to it. If Baker and Burge are correct, then there is no need to worry about formulating a metaphysical account that will accommodate the causal efficacy of mental properties conceived along antireductionist lines.

---

[29] See Baker's "Metaphysics and Mental Causation," in *Mental Causation* (1993) pp. 75-97; and Burge's "Mind-Body Causation and Explanatory Practice," in *Mental Causation* (1993) pp.

I will argue that Baker and Burge are wrong and that we need to provide a metaphysic of mental causation. I will then turn to what I consider the most promising solution to this problem, the trope solution, and explore its merits.

1. Causal Relevance without Efficacy: "Programming Relevance" and Counterfactual Dependency

In their "Functionalism and Broad Content," Frank Jackson and Philip Pettit[30] attempt to provide an account of the causal relevance of irreducibly mental properties. The sorts of properties they have in mind are the kind of higher-order properties typical of functionalist accounts of the mental. By a higher-order mental property, I mean here, a property that consists in having certain other properties that meet a given functional description is met. To take an uncontroversial non-mental case, consider fragility. This is a property that my glass has in virtue of having such-and-such a molecular structure. There may be many such structures that realize this property, but in general, something is fragile only if it has some structural property or other in virtue of which it breaks

---

97-121.

easily. There may be many structural properties that accomplish this, but something counts as fragile only if its underlying molecular structure does this.

For example, a fragile piece of glass will have a certain molecular structure that happens to meet the following functionalist description: With an input of slight force, we get an output of breaking. If we define fragility functionally, then we are concerned only with certain input-output relations a thing has--not how it may happen to realize these relations. But then any molecular configuration that realizes this input-output structure will have the higher-order property of fragility. Its lower-order properties will be its molecular properties. The higher-order property here is a property of those properties—the property of meeting the functionalist characterization at hand.

Or consider dormativity. A sleeping pill has this property in virtue of having such-and-such a range of chemical structures.[31] Again, there may be many ways in which this property is realized. But something counts as a

---

[30] In *Mind*, vol. xcvii, no. 387, July 1988. Henceforth J&P.
[31] And being in a world in which this structure has the effect of inducing sleep in normal human organisms.

dormative substance only if its chemical structure has the effect of inducing drowsiness in typical human organisms. This then would be a property of whatever chemical properties made something a soporific.

In the case of the mental, being a pain would be a higher-order property of whatever physical or neurobiological properties connect a body's inputs to its outputs in the right sorts of ways. If C-fiber firings do the work characteristic of pain--that is, of responding to tissue damage by leading to wincing, avoidance behavior, etc., then C-fiber firings count as pains for the functionalist.

But according to J&P, such higher-order properties do not enter into causal relations. For it is not a sleeping pill's dormativity that gets me to sleep--it is the chemical structure of the pill that makes me drowsy.[32]

Similarly, it is not fragility that is responsible for the glass breaking as it does. Fragility is a property that serves merely to underscore the fact *that* the thing in question will tend to break easily under the right

---

[32] Here they are in agreement with Ned Block's claim regarding the causal impotence of second order properties in his "Can Mind Change the World?"

circumstances. But it breaks so easily *because* it has one of the many molecular structures that are not very resistant to certain kinds of force. It is the underlying molecular structure, not the higher-order functional property that this structure has, that is thus causally responsible for the glass' breaking. For *it*, and not fragility, is what will come into contact with the various forces that cause the glass to break when it is dropped.

Carrying this analysis over to the case of the mental, it is not being a pain that causes anything--it is the underlying neurobiological properties that do the work. It is C-fiber firings that get me to wince, not the property these firings have of counting as pains.

J&P go on to distinguish between functional realizer states and role states, and this distinction is important for understanding the causal efficacy and relevance of the mental. Here is how they describe the distinction:

> Functionalism *per se* is a thesis about the truth-conditions of psychological attributions. It says that I am in mental state M when I have in me a state which is playing the causally intermediate

> role between inputs, outputs, and other states definitive of M. For short, I am in M only if I have in me a state playing the M role. The metaphysical question of which state *is* the psychological state is a further matter. Thus functionalism *per se* says that to be in pain is to have in one a state playing the pain role, where the pain role is defined in terms of being causally intermediate between inputs, outputs, and other states in one. And this means there are two candidates for the pain role: the state that plays the role, and the state of having the state that plays the role.

According to the functionalist story, we have a choice when it comes to identifying any given mental state. Consider, for example, a state of pain. Suppose C-Fiber firings realize this state in humans. Well, we can think of pain as the *having of this state*, which would make it a kind of "second-order" or "higher-order" state, or we can identify it with the "first-order" or "lower order" state that actually plays the pain role in a given body. The

former would be a role state in J&P's sense; the latter would be a realizer state. The realizer state is what realizes the mental property in question. The role state is the state of *having* a realizer of this property. This would be let's say, any organism's state of *having* C-fiber firings.

Let's try to make this distinction clearer. According to functionalism, anything that meets certain input-output specifications will count as a case of pain. So the property of being a pain will be the property of having the right sorts of causal connections with inputs, outputs, and other mental states, however realized. But since being a pain is an MR[33] property there will be many different lower-order properties that realize *being a pain* in different organisms. C-fibers firings will realize this property in humans, but M-fibers firings will realize it in Martians, and so on; and in each case we should expect that very different physical properties subserve the mental property in question.

Now, a human and a Martian will then both count as pain-feeling organisms if the foregoing is true. But we can

---

[33] Multiply realizable.

now ask: Where exactly is the state of pain in this story? Is it at the higher, functional level, which is shared by both the human and the Martian *insofar* as each is a pain-feeling system? Is pain, then, what is common to both the Martian and the human? If so, being in a state of pain would not be identified with being in a particular physical state, for this is different in the Martian case and the human case. If we identify pain as the thing that is shared by both the Martian and the human, this would make pain a "role state" in J&P's sense. It is the state of having some state or other which meets the functional characterization of pain. When I am in pain this is identical with my being in a state that has characteristic causal interactions with the environment, other mental states, etc. Hence my state of pain is my state of having something in me, *whatever it might be*, that performs the role characteristic of pain.

Notice the difference: the state of meeting the functionalist description of pain is the state of having C-fiber firings, because these happen to realize pain in humans. But having C-fiber firings is distinct from having the state that meets this account. For if C-fiber firings realized "feeling groovy" and not pain, then there could be

a state of C-fiber firings in me without my having the appropriate role state. For if there is nothing in me that realizes pain then even if C-Fiber firings occur within my body they do not add up to my having a state which plays the functional role distinctive of pain. I may have the state of *having C-fiber firings*, but because my C-fiber firings are not playing the right sort of role, this state would not count as the appropriate "role state."

We could also consider the realizers of pain as our candidate for being the mental state in question. Are the actual C-fiber firings that occur in my body pain? According to J&P, functionalists often obscure the issue here by gliding between two ways of taking about pain. It is not unusual to find talk of pain as an irreducible, MR phenomenon, thus implying that it is the higher-order, functional state that is shared across individuals and species that *is pain*.

But then one can also find talk about the *multiple realizations* of pain, the occupants of the causal role distinctive of pain, which suggests that the various realizers of pain are really the pains. After all, to be in a state of pain is to be in a state that meets a certain

functional description. But it is also to be in some particular physical state, at a particular time.

According to the first way of talking, being in a state of pain would be identified with being in a role state in J&P's sense. According to the second, it would be identified with being in a realizer state.[34]

J&P's suggestion is that it is the realizers and not the higher-order property or state in question that does the causal work:

> If we want psychological states to be causes of behavior, then we must regard them as realizer states. It is the state that fills the functional role, not the state of having a state that fills the role, which does the causing of behavior. And this looks like a powerful consideration in favor of identifying mental states with realizer states.

---

[34] Although it seems clear to me that if pains are identified with their realizers the functionalist has given up the antireductive component of the thesis.

Although they do not really give any arguments in the article under discussion for the claim that it is the realizer state, not the role state, which does the causal work, Frank Jackson does give an argument elsewhere for this claim.[35] In "Essentialism, Mental Properties and Causation," he writes:[36]

> Which of the two properties actually causes the breaking of a fragile object...when it is dropped--its fragility, or the categorical basis of its fragility?...The fragility of a glass is a matter of its having a nature that would cause the glass to break on dropping; but if the nature would do the causing, then, by Modus Ponens, it does do the causing...But the nature is the categorical basis. Ergo, the categorical basis, and not the disposition, causes the breaking.

When we consider the functionalist approach to defining mental states and properties, it amounts to

---

[35] In his "Essentialism, Mental Properties and Causation," in *Proceedings of the Aristotelian Society* 95, vol. 95 (1995) pp. 253-268.

defining a type of mental state *M* by considering the kinds of causal connections it has to the environment, other mental states, etc. But, to do this, there must be *something* in the creature in question that is doing all the causal pushing and pulling such that the functionalist characterization is met. This *something* would be the "categorical basis" of the mental state. Hence, something *counts as* fragile because of its molecular structure. But its counting as fragile does no causal work. The structure that realizes this property does it—*and that's why it counts as fragile.*

Hence, a C-fiber firing counts as a pain, only if it connects with a given body's inputs and outputs in the right sorts of ways. But then its counting as a pain is a consequence *of this C-fiber firing having causal connections* of various sorts. Realizers fit roles, *because of what they do.* Roles are just descriptions of what they do. But then, it is what fits the role that is doing all the work.[37]

This may look suspiciously like epiphenomenalism for

---

[36] *Ibid.*, p.256.

[37] Again, see Ned Block's "Can Mind Change the World?," for a good discussion of the causal inertness of higher-order,

the strict functionalist.[38] After all, if functionalism is true, then what it is to be pain cannot be exhausted by realizer states in J&P's sense. For realizer states only capture specific realizations of pain. They do not capture what it is to be pain in general. Hence being a pain, if irreducibly mental because functional in nature, would not have causal efficacy for J&P. Only realizers of pain would. This makes it look like all the causal work is going on at the physical level--the level of realizers. The mental level is itself a higher-order, causally impotent onlooker here.

Hence, mental properties like being a desire for beer make no causal contribution to the physical world. For it is the underlying neurological transactions in my body that actually get my muscles to expand and contract in the right sorts of ways such that I move toward the beer I want. But part of what makes my desire for beer the very desire that it is, is a fact about the world outside me. A state of beer-desire is one in which I am connected to beer in the right sorts of ways. The very same sorts of neurological states that get me to move in just this way might occur in a doppleganger of mine in a world in which Tweer, but not

---

or "second-order" properties like fragility and dormitivity.

beer, exists.

Let's imagine that Tweer is superficially identical to beer but chemically distinct from it.[39] Well, then in such a world my twin's body may gyrate in exactly the same ways mine does in my beer-filled world. But then our bodies would have moved as they did because of the underlying neurological state(s) occurring within us--not because of the desire states we are in. It is the realizer states, the neurological firings in our bodies, not the role states that happen to make our mental states the sorts of mental states they are, which move our muscles, bones, and limbs.

The intentional nature of my desire, which depends on my being related to a beer-filled environment in order to count as a beer-desire has no impact on how my body moves. But then I move, not because I want a beer, but because of how my neurons fire. For if I want a Tweer and the very same neural state realizes this desire and causes my body

---

[38] See Kim's *Mind in a Physical World* (1997) chapter 3.
[39] See Hilary Putnam's "The Meaning of Meaning," in K. Gunderson, ed. *Language, Mind and Knowledge, Minnesota Studies in the Philosophy of Science,* 7 (University of Minnesota Press: Minneapolis, 1975); and Tyler Burge's "Individualism and the Mental," in P. French, T Uehling, and H. Wettstein, eds. *Midwest Studies in Philosophy 4, Studies in Epistemology* (University of Minnesota Press: Minneapolis, 1979) for classic arguments defending the idea that content

to move in exactly the same ways, then it would seem that my desire qua desire plays no causal role.

This sounds like epiphenomenalism. For here, the desire-state, which is a role state in J&P's sense, is the broad state of my being in a certain relation to beer or Tweer facts in my environment. But these facts only serve to help classify my movements as instances of a desire state of a certain type—a beer desire vs. a Tweer desire. My movements are themselves caused by the neural firings in me, which are themselves blind to such extra-bodily facts.

Interestingly enough J&P think this is not a worrisome epiphenomenalism because they believe that the mental properties in question can serve to "program for," and hence prove causally relevant to, various effects that might be taken to have mental causes. Thus even if they are not causally relevant in the sense that they are *efficacious*, mental properties will serve as a kind of "guarantee" that certain effects will follow. This is supposed to show that these mental properties are not totally irrelevant to the effects they are expected to have from a folk psychological point of view. A mental property

---

is determined by states external to the body.

D may not cause *E*, but *D's* presence may ensure that *E* will occur. This is what J&P's program relevance says. It is an intriguing suggestion, and if they are correct it provides a solution to the mental causation problem as we have been considering it.

Here is an example to help illustrate their claim.[40] Consider hurricane Donald. Let's say it caused massive flooding in the streets. And suppose it did this in virtue of its physical properties. Let us suppose further, that the property of being a hurricane had no effect on a single drop of water that fell or a single gust of wind that blew.

But now suppose that hurricane Donald occurred with just a slightly different set of physical properties than it actually did. Suppose just one less drop of rain was involved in this counterfactual hurricane. Still, because Donald is a hurricane, the range of effects it will have is constrained in certain ways. It will not leave the streets dry, for example. Even though the physical facts do all the work here, knowing that Donald is a hurricane gives us explanatory power because being a hurricane "programs for" certain effects.

---

[40] This example comes from LePore and Loewer's "Mind

Or consider Hilary Putnam's example of a square peg not fitting into a round hole. As J&P point out, squareness as such does no causal work. All the causal interactions between a rigid square peg and the surface of a round hole into which it does not fit, involve the micro-particles of the objects in question. And the configuration of micro-particles that realizes one instance of squareness can be replaced by another, different configuration. But again, it would not be squareness as such that has done any causal work.

Yet the fact that something is square is not irrelevant. For if something is square, we know it will not be able to fit into a round hole.[41] In general, any rigid peg that is square will have this effect, even though it has it in virtue of its realizer state--the configuration of its micro-particles--and not its role state--its state of *being square*. But because the property of being square ensures that a rigid peg will not fit into a round hole, regardless of the details of its particular micro-

---

Matters" (1987).

[41] Of course, it must be square pegs, of a certain size we are talking about here. For a small enough square peg, could fit into a big enough round hole. But again, even these facts could be "programmed for" in J&P's sense. For this

configuration, it "programs for" the result that any square peg will be unable to fit that hole and others relevantly similar to it.

In the case of the mental, role states "program for" certain effects by indicating that a certain range of cause-effect relations are to be expected. If something is a pain, then regardless of its particular mode of realization, we should expect certain characteristic effects given the right sorts of input. Suppose my C-fibers are replaced by X-fibers, which are synthetically manufactured C-fiber substitutes with a chemical make-up different from ordinary C-fibers. Even though there will be a new, physically different realizer state in me, the range of effects one could expect from my being burned with a cigarette, cut with a knife, etc., will remain pretty much the same. In this sense, my role state, of having a state in me which realizes the state of being in pain, "programs for" various effects which are actually produced by the realizer state. Hence the mental property of being a pain, which is realized by my physical states, "programs for" various effects.

---

generalization is not realization-specific.

2. The Counterfactual Approach to Causal Relevance

Ernest LePore and Barry Loewer offer a similar account of causal relevance.[42] Though their concern is primarily with defending Davidson's AM,[43] I believe their approach is similar in important ways to that of Jackson and Pettit and is worth discussing here.

According to LePore and Loewer (herein "L&L") we can distinguish between two types of causal relevance: relevance$_1$ and relevance$_2$. For L&L, a property *F* of *c* is causally relevant$_1$ to *e's* being *G*, if *C's* having *F* "makes it the case" that *e* is *G*.[44] This will involve lawful connections between *F* and *G*. However, L&L do not think mental properties can have this kind of relevance since 1) there are no such nomic ties between the mental and the physical and 2) mental properties are distinct from physical properties. Rather, mental properties can have a weaker, counterfactually based relevance according to which the following claims are true:[45]

---

[42] In their "Mind Matters," in *The Journal of Philosophy*, 1987, pp. 630-642.
[43] Anomalous monism.
[44] *Ibid*, p. 635.
[45] *Ibid*, from p. 635.

i. *c* caused *e*.

ii. *Fc* and *Ge*.

iii. *~Fc > ~Ge*.

iv. *Fc* and Ge are logically and metaphysically independent.

This is what relevance$_2$ amounts to. The idea here is that a property may be causally relevant$_2$, even if it is not efficacious with respect to a given effect, if the non-occurrence of the property would guarantee non-occurrence of the effect having the property it did have. Consider for example my wanting to raise my hand. If I had not wanted to raise my hand, then all other things being equal, my hand would not have moved as it did. L&L describe the situation in terms of possible worlds. So, in a possible world, most like the actual one at time T, in which I do not want to raise my hand, my hand does not move.

Where J&P offer a notion of relevance which assures a *range* of effects given the presence of a mental property M, L&L claim that a mental property M is relevant in a stronger sense: had it not occurred a given effect would

not have occurred. Yet like J&P, they concede that all the causal work is being done at the physical level.

So suppose I desire to drink some water and extend my hand. It may be true that being a desire played no causal role here--it is the neurological properties that realize my desire that get my arm to move as it does. But, according to L&L, if the property of being a desire had not occurred here, then I would not have extended my hand. In a possible world most like the actual one at time T in which I do not desire a glass of water, my hand does not extend. This is true even if there are no laws between mental properties and behavioral effects. What grounds the causal relation here are physical, presumably neural properties, and the nomic connections they have to physical, behavioral properties. But despite this, L&L think that because certain counterfactuals would be true in the sorts of cases at dispute in the mental causation debate, this makes mental properties relevant in some interesting sense.

3. Problems with Causal Relevance without Efficacy

Both J&P and L&L think that causal relevance without causal efficacy can save mental causation. But is this a

promising approach to the problem? Before answering this question, let us consider what T. Huxley, the famous epiphenomenalist said about the causal relevance of the mental:

> Our mental conditions are simply the symbols in consciousness of the changes which take place automatically in the organism;...to take an extreme illustration, the feeling we call volition is not the cause of a voluntary act, but the symbol of that state of the brain which is the immediate cause of the act. We are conscious automata.[46]

On Huxley's view, mental states do not bring about behavioral consequences. Hence, my desire to raise my arm only seems to move my arm because the alleged cause and effect here are actually effects of a common cause.

One thing that may seem troubling about J&P's "program relevance" and L&L's counterfactual dependence approach is their resemblance to Huxley's epiphenomenalist account of

---

[46] T.H. Huxley, "On the Hypothesis that Animals are Automata," in *The Philosophy of Mind* ed. Brian Beakley and Peter Ludlow (M.I.T. Press: Cambridge, 1992) p.136.

the mental-physical relation. On Huxley's view, pains, desires, etc. are causally impotent effects of brain states. They are "symbols in consciousness" which do no real causal work. My pain may serve as an "indication" that I have suffered tissue damage; but it does not in any way bring about my screaming "Ouch!" But since pains are effects of the underlying neurological states that would be expected to bring about winces, cries, whimpers, etc., then pains, even though they are not causally efficacious, would "program for" the alleged effects of pain.

Similarly, in any possible world most like the actual world in which I did not feel a pain at time T, my screaming "Ouch!" would not have occurred. Only these effects would not be products of pain states. Rather they would be the result of whatever neurological cause(s) gave rise to the pain.

So if pain, along with wincing, whimpering, crying etc., is an effect of a common neurological cause, this would do little to comfort supporters of folk psychological claims like: John winced, *because* he was burned, where 'because' is understood in a causal sense. If John's wincing is caused by his C-Fiber firings, and not his pain,

then he did not wince *because* of his pain, even if that pain "programmed for" his wincing or even if, had he not experienced pain at that moment, he would not have winced.

Of course, if John's pain is identical with his C-fiber firings, with his realizer state, then there should be no epiphenomenon problem. Consequently, J&P could retort here by pointing out that their case is importantly different from the epiphenomenalist case outlined above. For it is still open to them to identify pains with realizer states, thereby securing the causal efficacy of these entities. Indeed, this is precisely what J&P suggest in their article; and as advocates of AM, L&L claim that mental events are physical events so this reply is open to them too. If pains are C-fiber firings, pains do whatever causing C-fibers firings do.

The problem however is that we are trying to secure the causal efficacy of mental *properties*. There is little disagreement that mental states, since they are token identical with physical states, will enjoy whatever efficacy those physical states enjoy. The problem is whether mental *properties* do any work. The problem, as you will recall, is that if mental properties are distinct from

physical properties and only the former may have causal commerce within a causally closed world, then the latter do all the work.

So mental properties are higher-order properties which merely "program" for their effects for J&P. For the functionalist does not want to identify pains with any specific realization of that state. Remember, functionalism, as it is being considered here, is a form of antireductionism. Being a pain is an irreducibly mental property--a higher-order property realized in myriad ways across physically diverse systems. But on J&P's view, "role states" and hence higher-order functional properties like being an M, are mere accompaniments to the causal scene. They are "relevant" in the weak sense that they serve as an indication that a given range of effects is in the offing. But Huxley could also describe the causal relevance of the mental in this way.

And the same goes for L&L's version of AM. For them, mental properties cannot be reduced to physical properties and may be realized in all sorts of physically diverse ways. Hence, they are not themselves allowed to make a causal entry into the physical world. L&L's

counterfactually grounded "mental causation" allows for physical properties to do all the causal work while the mental properties serve as mere "indicators" of the effects of physical properties.

To see this point better, consider the non-mental example Robert Van Gulick[47] gives in criticism of L&L's view. Suppose there is a chemical compound TTT that when successively heated and cooled goes into molecular configuration state BAM. When TTT is in state BAM it is explosive. But suppose that when and only when TTT is in BAM it also turns blue. Intuitively, we would want to say that when a match is dropped into a glass of blue BAM TTT the TTT's being blue has no relevance to the resulting explosion. Indeed, this would seem like a perfectly straightforward example of non-mental epiphenomenalism. Yet if a given sample of TTT is not blue, then it won't explode. Hence, a la L&L, if this TTT were not blue, it would not have exploded. We could also say that being blue "programs for" TTT's being explosive. Yet would we want to say that mental properties are relevant in *this* sense?

---

[47] In his "Who's in Charge Here? And Who's Doing All the Work," in *Mental Causation* ed. by John Heil and Alfred Mele (1993) pp. 233-256.

If so, then how have we escaped the charge that anti-reductionism is committed to the causal impotence of mental properties? It would seem that such properties are without *any efficacy at all* on this view.

If a state's being mental merely "programs for its effect," or supports counterfactuals as L&L claim, then these mental properties play no more causal role than would a purely epiphenomenal accompaniment of a brain state. *Perhaps* they do more than this, and much of our folk psychological talk implies as much, but there seems nothing in the accounts we have considered, other than faith in the truth of these folk psychological explanations, to give us good reason to think that this is possible. Hence we may still worry--does the mental cause anything insofar as it is mental? Is relevance in a sense stronger than J&P's epiphenomenalist "program relevance" and L&L's counterfactual dependence approach possible?

4. Explaining Away the Metaphysics of Mental Causation

If what we think and feel does indeed have no causal relevance to what we do, then as Huxley observed we have

good reason to conclude that we are automata. But this possibility seems so counterintuitive from a common sense standpoint that one might argue that whatever premises lead to this conclusion should simply be rejected. Lynne Rudder Baker[48] and Tyler Burge[49] have both taken this strategy, arguing that the conclusion of epiphenomenalism should be thought of as a reduction to absurdity of the metaphysical premises that lead to it. What are those metaphysical premises? For Baker it is the causal closure of the physical and for Burge it is the idea that mentalistic explanation must involve physical causation. In this chapter, I will limit my discussion to Baker's argument. What I say in criticism of Baker's view can be applied *mutatis mutandis* to Burge's view. For both argue that we ought to give up trying to find a metaphysical account that makes sense of mental causation and physicalism at the same time. It is this move of which I am critical.

In her fascinating paper, Baker argues that we have more confidence in the success of mentalistic explanations than we do in the basic tenet of physicalism according to

---

[48] In her "Metaphysics and Mental Causation," in *Mental Causation* (1993) pp. 75-97.
[49] In his "Mind-Body Causation and Explanatory Practice," in

which all causation involves physical events and properties as causes. By mentalistic explanation, Baker means typical folk psychological statements that refer to mental states as causes of behavior, like "Sally returned to the living room *because* she thought her keys where there and she wanted to find them." Since we have already outlined the premises that are thought to lead to the problem of mental causation, and Baker does not offer a different account of how the problem arises, we needn't rehearse that argument here. What is interesting is Baker's solution to or, better "dissolution" of, the problem. And that involves singling out one of the crucial premises: the causal closure principle (herein CCP) and the notion of causation normally tied to it.

According to Baker, this principle is deeply problematic because it not only seems to undercut the possibility of mental causation, but it threatens the causal efficacy of all irreducibly non-physical properties. This means that any properties of the so-called "special sciences" are in equal danger of being made epiphenomenal if the causal closure principle is taken seriously.

Mental Causation (1993) pp. 97-121.

So her solution is to re-think the CCP and the notion of causation that generates the problem of accounting for the causal efficacy of non-physical properties. She writes:

> My proposal, then, is to dismantle the problem of mental causation by rejecting the metaphysical background picture that generates it. If we accept paradigm cases of explanation in the sciences and in everyday life, and if we take the notion of explanation to be prior to that of causation, then the idea of a 'complete cause' in CCP hardly makes sense.

And furthermore that:

> Systematic explanatory success, in either science or everyday life stands in no need of metaphysical underpinning.[50]

The idea is to put aside worries about the causal efficacy of non-physical properties by uprooting the assumption that

---

[50] *Op.Cit.*, p.94

only physical properties can have a causal impact on the physical world. Geological, biological, meteorological, psychological properties, and so on, do figure into explanations that seem to rely on causal relations between them and physical events and properties. The idea that the most basic physical properties might somehow "gobble up" all causal efficacy of the macro-level, that they provide the "complete cause" of physical effects, seems to undermine common sense and scientific practice.

Hence, according to Baker, we should not think that there is any problem with mental causation in particular because our explanatory practices provide stronger confirmation of its reality than the claim that all causation involves physical properties. For as she points out, we don't know much about the most basic physical properties of the world. We know a lot more about macro-properties and their relations. Our insistence then that the bottom level provides the "font of all causality"[51] makes it seem as if we have betrayed common sense in favor of a rather obscure commitment to causality as an

---

[51] Baker, p.93 in *Mental Causation*.

"objective relation in nature"[52]--one distinct from our explanatory practices and epistemology.

As Kim points out,[53] there is a major oversight here. The assumption that psychological explanation, like much scientific explanation, is causal in nature was itself a source of heated debate in philosophy in the 1960's. It was Donald Davidson who managed to convince a majority of philosophers that reason-giving explanations are a form of causal explanation in his landmark 1963 paper "Actions, Reason, and Causes."[54] But that argument involved a theory of causation, events, and explanation. This shows that taking explanation as our starting point does, itself, require various metaphysical commitments. Why should we assume, with Baker, that such folk psychological explanations are causal? If we do, it seems we have already presupposed a lot of metaphysics. The problem of mental causation can be seen as the attempt to sort out those assumptions to help understand just what sort of "metaphysical underpinning" we have available.

---

[52] *Ibid.*
[53] In his *Mind in a Physical World*, p.63.
[54] Reprinted in *Essays on Actions and Events.*

Perhaps we feel that there are good reasons to assume that our mental lives causally impact on the physical world. Perhaps we think that unless what we think and feel has a causal impact on what we do, then we are not agents--that it is just serendipity when my arm moves at just the time I want it to--and that this is absurd. It seems to me that, barring eliminativists, few would disagree that we are agents and that our mental lives do affect causally the physical world.

But this just shows that we need to provide an account which both supports our commitment to mental causation and the metaphysics behind it. The reason for this is that for any given instance of behavior, we can expect a causal story that makes reference only to neurophysiological phenomena. And the success of neurophysiology would, on Baker's argument, provide good reason to take it seriously as a domain of legitimate causal explanation.

Now, if we insist further and quite sensibly, that neurophysiological explanations are distinct from intentional explanations, we are left with a mystery. How do they relate to one another? It is very tempting to make the following eliminativist argument: since all the causal

linkages here are purely neurophysiological in nature, any alleged "mental causes" are unnecessary and hence should be sliced off with Ockham's razor.

Of course Baker would point out that such a move relies on singling out one level of description as the "real" explanatory level, unjustly leaving the other out in the netherworld of pseudo-explanation. Perhaps we do not need to view this as an exclusive choice. For it is possible to have different descriptions of the same phenomena. Indeed, this is what Davidson has famously argued in his AM. Mentalistic descriptions, and hence intentional explanations, refer to the very same phenomena picked out by physical descriptions. Mental explanations and neurophysiological explanations then are not in competition, but are rather alternative modes of picking out the very same patterns of the world around us. But what shows that these explanations are not in competition for Davidson is the claim that mental entities *just are* physical entities and that causation is extensional in nature, that is, that how we describe things has no impact on their causal efficacy.

The difference between mental and physical explanations has to do with how mental states are picked out--whether they are described as states in which a rational being ought to be, or they are picked out as states of a human being's brain, with no implicit assessments of rationality. The bond between the mental and the physical is forged in the form of a token identity thesis. This is one of Davidson's "metaphysical underpinnings" for the reality of mental causation.

Whether this story is ultimately satisfying is a question I will take up in chapter five. But it is worth noting that, at the very least, it does answer the question of how intentional and physical explanations relate to one another: they are different ways of looking at the same phenomena, for mental events *are* brain events on AM. Hence they do not refer to distinct states which may be thought to be in competition for causal roles. Hence neurophysiological and intentional explanations may both be true of the same explananda. We can have a metaphysical underpinning of our explanatory practices through identification after all.

However, if Baker's account is right, we are left with a gap in our explanatory practices. We still need an answer to this question: What do neurophysiological explanations have to do with psychological explanations? There is good reason to think that there is a strong relation between the mental and the physical, as Baker herself concedes in her acceptance of some form of mind-body supervenience. After all, we know that what happens to the brain has effects on our mental lives. This is why psychotropic drugs are possible, for example. But if we leave a sharp gap between intentional explanations and physical, causal explanations we are left with a mystery. Why does anything that happens to the brain have *any* effect on the mind? Given the fact that mental explanations and neurophysiological explanations have proven themselves successful at picking out causal relations, how do such explanations relate to one another? If we abandon the search for a metaphysical underpinning, as Baker urges, we must embrace these mysteries.

Of course most philosophers today answer the sorts of questions raised in the last paragraph by pointing to some form of physicalism. Indeed, Baker seems to do this much in

accepting a version of the supervenience thesis. But this leaves her with the philosophical task of showing just how our mental lives relate to the physical world in more metaphysical detail. To simply insist that folk psychological explanations are enough to account for mental causation, without a metaphysical theory to support this claim, seems rather cavalier if not downright obstinate.

Why not revert to a straightforward dualism of mind and body? Baker's own reliance on supervenience indicates why: we tend to think that a departure from the causal linkages of the physical world would invite hopeless mysteries. But if we are going to accept this much of the physicalist hypothesis, it seems odd to stop here and insist that intentional explanations are causal explanations, even if we can't show *how* this is true.

I think Baker would point out here that she does think that there is some way to show how such alleged causal explanations are true, and that we are not left with a total mystery here after all. She writes:

> [f]or example, when Jill returns to the bookstore to retrieve her keys, what she thinks is that she

> left her keys on the counter and that she wants them back. What she thinks affects what she does in virtue of the following explanatory fact: if she hadn't thought that she had left her keys, then, other things being equal, she wouldn't have returned to the bookstore...

Unfortunately, as we have already seen, the appeal to counterfactuals as a grounding for mental causation gives us only a very weak form of causal relevance—one compatible with epiphenomenalism. Of course Baker would not be moved by this concern because she would insist that the success of our explanatory practices is enough to ensure that any metaphysical assumptions that lead to an epiphenomenalist conclusion must be wrong. But then why try to account for the phenomenon at all? Why try to show that it has any metaphysical basis through counterfactual claims, supervenience, or any other accounts?

Part of the problem here is that Baker at least implicitly recognizes the need to provide an account of the mind-body relation consistent with the demands of mental causation. Supervenience is supposed to do this. And the

success of our explanatory practices is supposed to put to rest any worries that the subvenient physical basis of the mental-physical relation does all the causal work.

So Baker recognizes the need to make the mind physical in some sense to explain the apparent causal connections between physical and mental events that are implied by much of our causal-explanatory talk. Since she also finds that this move threatens our commitment to mental causation, by making the physical entities involved do all the causal work, she should recognize that this means that we have our philosophical work cut out for us. To insist that mental causation is real and that we can get by without knowing how physical explanations and mental explanations relate to one another seems to me an unsatisfying response to the problem. It is to trade in the insolubility of the mental causation problem for a kind of mysticism concerning the ways in which our mental lives are connected to the physical world. It does not dissolve the problem of mental causation—it shrouds the reality of mental causation in a mysterious relation of mind-body supervenience that is supposed to somehow both threaten and support our causal explanatory talk. It threatens by suggesting that all the causal work is

done by the subvenient physical properties upon which the mental supervenes, and "supports" by showing that the mental is intimately connected to the physical after all, in so supervening. But it seems that if we want to elucidate this relation and vindicate mental causation, we need to understand better how the mind and body interact--for this *is* a metaphysical problem.

5. David Robb's Trope Approach

In his "The Properties of Mental Causation,"[55] David Robb puts forward the basis for an ingenious solution[56] to the problem of mental causation.[57] Robb's solution is simple:

---

[55] David Robb, "The Properties of Mental Causation," in *American Philosophical Quarterly*, Vol. 47, No. 187, April 1997

[56] There have been at least a few versions of this solution put forth in the literature c.f., Robb, David. "The Properties of Mental Causation," in *American Philosophical Quarterly*, Vol. 47, No. 187, April 1997. See also John Heil's *The Nature of True Minds*; and Cynthia and Graham MacDonald's "Mental Causes and Explanation of Action," in *Mind, Causation and Action*, ed. L. Steven, R. Squires, and J. Haldane (Basil Blackwell: Oxford, 1986); as well as their "How to be Psychologically Relevant," in *Philosophy of Psychology* (Oxford: Blackwell, 1995).

[57] This theory has also received some virulent criticism By Paul Noordhof in his "Do Tropes Resolve the Problem of Mental Causation?" in *American Philosophical Quarterly*,

identify mental tropes with physical tropes. This eliminates worries about causal closure, for now mental property instances just are physical property instances. But doesn't this entail rejecting the *distinctness* of mental properties? No, it does not. For mental *types* may still be distinct from physical *types*--they may merely supervene on them. Being a pain in general may not be reducible to any determinate physical property. But every instance of being a pain will be identical with an instance of *some* physical property or other, regardless of how heterogeneous the total set of such physical properties might be. This makes for a property identification that is consistent with the multiple realizability of the mental.

Suppose Robb is right and mental property tropes are identical with physical property tropes. The problem this is supposed to solve is that of securing the causal relevance of the mental. But this raises another worry. For any given mental trope will have both a physical aspect and a mental aspect since each such trope will be both an instance of a mental property and a physical property. We now face the

---

1998, pp. 221-228; and in Steven Yablo's "Mental Causation," in *The Philosophical Review*, 101 (1992), pp. 245-80. See also Douglas Ehring's "Mental Causation, Determinables, and

problem of determining which of these aspects is causally relevant to any given effect *e*. That is, we need to account for the following fact:

The Principle of Trope "Quasation:"[58]

If Trope *T* has mental aspect(s) *M* and physical aspect(s) *P*, then when *T* brings about effect *E*, we can sensibly ask whether *T* brought about *E qua* M *or qua* P.

This principle states that whenever a property instance causes some effect, we need to determine which of *its* aspects were causally relevant in bringing about the effect in question. When I crush a grape with my foot, the cause here may be a trope of a physical property: it will be a trope of being a physical movement of some kind. But it will also be an instance of a mental property insofar as it is an instance of being an intentional action.

Even if this instance of being an intentional action is identical with this instance of being a physical action of

---

Property Instances," in *Nous*, 30, 1996, pp.461-80.

[58] This locution comes from T. Horgan (1989) and is intended to capture to meaning of "causing qua." When *C* brings about

some kind we may sensibly ask: In virtue of which feature did the cause bring about its effect? Did the trope that is identical with the physical property in question here do its causal work without its mental aspect playing any role? If so, it would seem to make that mental fact irrelevant from a causal point of view.

Hence, like Davidson's AM, according to which every mental event has both a mental aspect and a physical aspect, the question of the causal relevance of *trope aspects* arises for the trope theorist. For we may now ask the following question: Which of a mental trope's aspects is causally relevant to its effect--its being a mental trope, or its being a physical trope? This is essentially the question that Paul Noordhof thinks spells disaster for a trope approach of the sort argued for by Robb.[59]

According to Robb, however, this question should not be a worry for the trope theorist because the question should not arise in the first place. The idea is not so much that properties do whatever causal work they do *simpliciter--it is that properties do not have various aspects that may compete for causal relevance at all*. Hence, there is no sense to the

---

and effect *E* qua *P*, we may say that *C "quased" E*.

question whether a given property F caused anything qua F or qua G or what have you. Contrasting this alleged problem with that of the apparent causal competition between the mental and physical features of events on Davidson's AM he writes:

> But we cannot use the same considerations to raise the *qua* problem for causally relevant properties, whether or not we view those properties as tropes. A causally relevant property F simply does not have various aspects such that one can legitimately ask whether some but not others are responsible for F's being causally relevant.[60]

He then gives an example in defense of this claim:

> Suppose, for example, that a red ball dropped on a sheet of metal causes a dent *qua* massive thing, but not *qua* red thing. It would then be odd to

---

[59] Paul Noordhof, 1998.
[60] David Robb, 1997, p.191

> object: 'Yes, perhaps its mass is causally relevant here; but is it causally relevant *qua* mass? Surely a good answer here is: 'Its mass is not relevant *qua* this or that, it is just causally relevant, period.'

But as Noordhof points out, we can raise "the qua problem" at the level of mental/physical tropes in the following way.

We must recognize that when *C* brings about *E* some features of *C* will be relevant to bringing about *E*, whereas others will not. It may be sensible to ask, for example, whether the fact that a given trope of being a pain occurred on a Tuesday had any causal relevance to my wincing. After all, this may be an aspect of the trope, and it is conceivable that one might wonder whether it had any causal relevance to the effect in question.

Similarly, the trope may be mentioned in a thesis on mental causation. This would be an aspect of the trope as well. Did it have any causal relevance to my screaming "Ouch!"? Given these examples, it seems clear that it is not the case that properties are "aspectless" in the way

that Robb suggests.[61] If it makes sense to deny the causal relevance of a given instance of painfulness being referred to in a doctoral dissertation, it makes sense to think of a trope as having various aspects. For irrelevant aspects are, nonetheless, aspects. Clearly, a given trope may exist within a given stretch of space-time, with relations to entities that make up that slice of space-time. And any of these may be properly called aspects of the trope in question.[62]

But here we must be aware of the difference between *causal efficacy* and *causal relevance*. For if Ausonio Marras is correct, the slide between these two notions has obscured much thinking on the mental causation problem.

In his "Non Reductive Materialism and the Problem of Mental Causation," Marras argues that causal relevance and

---

[61] He describes them as such on p.191 of his "The Properties of Mental Causation," (1997).

[62] Of course, if I mention a pain trope in a dissertation on mental causation, this shows that the trope may have an aspect that exists long after the trope itself has ceased to exist. This may seem strange, but no stranger than the fact that my great grandfather acquired the property of being my great grandfather only long after he has ceased to exist. We are talking about clearly extrinsic properties here. And once we realize that causal relations are *relations* and as such will involve many extrinsic aspects of the causal relata involved, I don't think we can simply dismiss extrinsic aspects as less than genuine or less than fully

causal efficacy are importantly distinct in that the former is essentially ontological, whereas as the latter is epistemological in nature.[63] The idea behind Marras's claim is as follows. When a given cause/effect relation holds it involves particular tokens. When *C* causes *E*, *C* and *E* may both fall under types *M* and *P* respectively. But their falling under types does not in any way add to, or diminish, the causal powers that *C* may have with respect to *E*. That *C* falls under *M* may be informative, but clearly it is not involved in the pushing and pulling whereby *C* brings about *E*.

But causal explanations will often be description-selective. They will involve picking out the events involved under descriptions that enable us to better understand why, or how, *C* brought about *E*. This may involve describing *C* in a way that allows us to subsume it under a causal law whereby it can be seen to be connected to the effect *E*. Again, it is not the mental event that is being denied causal efficacy here, it is a certain description of

---

real aspects.

[63] In his "Nonreductive Materialism and Mental Causation," in *Canadian Journal of Philosophy*, vol. 24, number 3.

that event--one of its classificatory properties--that is being denied *causal relevance*.

Although his concern is with events, I think Marras' reasoning can be applied to the case of tropes and the worry raised by Noordhof. For Nordhoof's complaint is that a mental trope's *being mental* plays no causal role with respect to any of its effects. But that is just a point about how the entity in question is being type-identified. Why should such type identifications be expected to play any causal roles in the first place? As Frank Jackson points out, the expectation that our descriptions of things will somehow affect their causal powers betrays a belief in "word magic."[64] But why should we believe in word magic?

As Marras points out, such descriptions are "abstract entities" and abstract entities don't cause changes in the spatio-temporal world. "It is only concrete exemplifications that do such causing."[65] But when we consider causal explanations, such classificatory designations may or may not be informative. When I am told that the event that was featured on the front page of

---

[64] Frank Jackson, in his *Mental Causation* in *Mind*, vol. 105, no. 419, July 1996, p.387.
[65] Marras, p.470.

Tuesday's *Times* killed innocent people, I may be told a true causal statement. But this statement may have poor explanatory value because that classification of the event does not reveal a regularity that would normally explain the sort of causal relation in question here. It was the hurricane, *which happened to be featured on the front page of the Tuesday Times,* that killed innocent people. How this event is typed seems irrelevant from a causal point of view. For if causation is extensional, then how causally efficacious entities are typed should not touch the question of whether they were efficacious at all.

But making sense of events in the world will involve describing them in certain ways, many of which will not make clear their instantiating causal regularities of various kinds. Here is how Marras makes the point:

> Suppose that, in fact, c caused e, where 'c' and 'e' are names or rigid designators of occurrent, token events. To explain what caused e, it is not enough to 'ostend' or 'display' its cause by mentioning c; rather, we need to convey the fact that c's causing e is an instance of a law. It is

> surely not a law that c caused e [because they are particular events]…The law will have the form: 'events of type F (of which c was a token) cause events of type G (of which e was a token)'. What displays the proper form of a causal explanation, therefore, is not 'c explains e,' but rather 'c's being F…explains e's being G.'

Particulars may be involved in causal transactions but causal explanations succeed by showing how these particulars can be subsumed under causal laws. But causal laws relate *types*, not tokens, to one another. None of the particulars involved in such transactions are themselves in causal laws. Why not? Because laws are *abstract entities* and tokens are concrete particulars.

So it is only certain of these tokens' *descriptions* that will be mentioned in various causal laws—namely, those that refer to them as being instances of types that do, in fact, figure into causal regularities. But the entities themselves do not appear in such laws. C-fibers do not appear in laws—'C-fibers' do so.

At the same time, the descriptions under which the entities in question can be seen to be instances of types that do figure into nomic regularities have no causal impact on the world. For they are, as Marras points out, abstract entities.

Part of the problem with the trope approach as presented by Robb is that it does not make this fact clear enough. The advantage of the trope move is that it limits our attention to concrete exemplifications of property types and universals thus meeting Marras's sensible requirement that causal relata be particulars, not types or universals. Whereas we would expect to find the former in causal relations, we might expect to find the latter only in causal laws. For it is not being a pain in general that causes anything--but a particular instance of being a pain, which occurs at a particular time, which does. But such exemplifications are themselves totally particular and would thus seem to be fit for causal work.

Unfortunately, Robb has focused on exemplifications of second-order classificatory properties like *being a pain*. So an instance of *being a pain* is identical with an instance of *being a C-fiber firing* let's say. This would

seem to guarantee the efficacy of the trope in question by identifying it with a physical trope. So far so good.

But if what Marras has argued is correct, and I think it is, there is a confusion here. Being an instance of being a C-fiber firing will be no more causally efficacious than being an instance of a Cartesian ectoplasm "firing." This is not because being a C-fiber firing involves a commitment to Cartesian dualism. It is rather because *being a C-fiber firing* is an abstract entity. It is a type individuation of a particular token. So *being a pain* does not seem to get any causal efficacy here if it is thought to inherit that efficacy from the physical trope with which it is identified since that trope, as an abstract entity, does no work either. In the case of type properties and their mental tropes, inheritance is not all that it is cracked up to be.

Understood this way, causal relevance is a fact about types, and causal efficacy is a fact about tokens. Hence if we can show that mental types are causally relevant, perhaps we will have done a better job of accounting for our folk psychological intuitions. For from that vantage-point, it is generally more informative to learn that

Jones's arm extended because of his desire to pick up the glass in front of him than to learn that he did it because of some neurological transaction between his brain and the muscles in his arm. And if functionalism is right, that desire could have been realized by some neurological configuration other than the one it actually had on a particular occasion. So we might think that the vindication folk psychology requires showing John's desire to grab a beer to be more causally relevant *as a desire* than as a neurological event of a certain sort. Otherwise, his extending his arm had little to do with the fact that he *wanted* a beer. Is there a way to account for this folk psychological intuition?

In what follows I will claim that the problem underlying the issue of mental causation, as it as been formulated here, is rooted in the notion of reduction that is at stake for the antireductionist. Once we understand just what reduction entails, I think worries about causal relevance for the trope theorist will be seen to be unreasonable. For the question of reducibility does not affect the issue of causation one way or the other, contrary to what we, and the participants in the debate,

have been assuming. If tropes of non-classificatory mental properties are identical with physical tropes, we will have the best possible basis for a metaphysic of mental causation. By non-classificatory properties I mean here any properties that are not merely taxonomical. The property of being a mental entity of some kind/type is a classificatory mental property. It is the property of falling under a particular mental type. For example, if the functionalist is correct, then being a pain is defined in terms of being a state that plays a certain kind of causal role. Any states that play the specified causal role thus count as pains and may therefore be said to have the property of being a pain. In this sense we can think of functionalism as providing a semantic thesis about general classificatory mental predicates.[66] For functionalism tells us what must be the case for predicates like "x is a pain," "x is a fear," "x is an itch," "x is a belief that p," etc. to be satisfied. In general, classificatory mental properties will be properties described by the locution "being an x,"

---

[66] Davidson does something similar in outlining the general conditions necessary for intentional state attribution, though he does not give an account of how to ascribe specific states with specific contents. He only undertakes the general conditions of such a procedure.

where x refers to some type of mental state/event, like desire, belief, pain, tickle, etc.

So, if something is a pain, then we know that there is *some* physical state or other, and hence, some physical property or properties, which will be expected to do the causal work characteristic of pain. But any such lower-level physical property will also be a mental property of a non-classificatory sort. For it will either be a phenomenal or intentional property of some kind. It will not be the property of *being such a mental entity*, for that is what the second order functional property is, but it will be the qualitative feel or intentionality of a particular mental state/event.

So if we consider a pain, its property of being a pain will be a mental classificatory property. Its non-classificatory properties will be all those properties that are not properties the pain has in virtue of falling under a given type. There are many such properties. In particular though, I have in mind the qualitative features of mental states. These are the sorts of mental properties we need to get into the causal mix in order to account for the common

sense notion that, for example, pains cause us to scream because they hurt.

Painfulness is a mental property in the broad sense that it is a qualitative feature of mental states. But painfulness does not depend on being characterized in any particular way in order to have the sorts of causal powers common sense might demand of instances of painfulness, i.e. that they cause wincing, crying, avoidance behavior, distress, etc. Such characterizations should not be expected to play any causal role, as we have seen. So common sense demands that a pain will cause me to flinch, not because it is a state with the property of falling under the type *pain*, but because it hurts. It is the painfulness of a pain sensation, not its counting as a pain for which the trope approach must account. But the trope move can accomplish this if painfulness tropes are identical with tropes of physical properties. And this goes for the tickle-ishness of tickles, the itchiness of mosquito bites, and so on. And this has nothing to do with the question of whether the characterizations of mental states, and the resulting classificatory properties of such

states, enter the causal scene. We needn't worry about word magic after all.

This move then shifts attention to property *instances*, or what I have been calling tropes. It is those particular instances that we need to incorporate into the causal scene. But, I will remain neutral on the ontological status of classificatory mental universals like being a pain, being a belief that p, being an itch, etc. My point is not that they do not exist, or are simply linguistic constructions. It is simply that we have good reason not to expect these sorts of properties to do causal work. Why? Because it is the properties in virtue of which a state satisfies the classificatory predicate in question, like being a pain, that should be expected to do causal work. One obvious reason for this claim is that the functionalist predicates in question are defined in terms of causal capacities. It is then the prior causal capacities of states like C-fiber firings that qualify them for the status of painhood. But then the fact that such status is earned is not itself something that should be expected to do causal work. Rather it should be thought of as an *indication* that some other causally efficacious property or

set of properties is present. Hence that something satisfies the predicate "being a pain" is a consequence of its causal capacities, all of which do their causal work in complete independence of this classificatory fact. The properties of a state, like a C-fiber firing, in virtue of which it satisfies the predicate "is a pain" are those properties that get an agent to wince, scream, engage in avoidance behavior, etc. under the right sorts of circumstances. If some of these properties are mental in the sense that they are qualitative or phenomenal then, as long as they are identical with some physical property tropes, we will have made the best possible defense of non-reductive physicalism as an account of mental causation. Instances of non-classificatory properties, i.e. instances of itchiness, painfulness, ticklishness, etc., if identical with instances of physical properties (i.e. physical tropes) should be able to do whatever causal work one could reasonably demand of the mental.

Unfortunately, I do not think this strategy will work for intentional properties. When we come to Davidson's AM we will see exactly why. Part of the problem is that attributions of intentional properties seem to depend on

facts outside of agents' bodies, such as the contents of their environments and the attributions made from the perspective of interpreters of those agents. But if we think that all bodily movements have causes that are internal to the bodies of agents, i.e. neurophysiological causes, then such external factors would not seem to be plausibly construed as the properties involved in the movements of one's muscles, bones, and limbs that are involved in action. These can be accounted for by neurophysiological transactions within an agent's body.

I think Robb's basic strategy is sound after all, even if it is limited to the case of phenomenal properties. But to see how it can do the work it is supposed to do, we must examine the notion of reduction that has informed this problem and see whether it really justifies the misgivings concerning the possibility of mental causation that we have been considering.

Chapter III: Reduction

I would like now to turn to the question of reduction by looking more closely at the functionalist multiple realizability argument and its claim that mental properties are irreducible. I will argue that the irreducibility thesis of functionalism concerns general classificatory predicates/properties and that the functionalist must allow for the tropes of mental experience to be identical with various physical tropes. But this is consistent with the realization relation anyway.

What is at stake is whether the realization relation will preserve neat isomorphisms between psychological kinds and physical kinds or between psychology and some physical theory generally. Multiple realization says this will be

quite unlikely, as a matter of empirical fact, and so there will be no reduction between mental kinds and physical kinds. But the mental tropes of experience will be identical with physical tropes and hence, there should be no special worries about their causal efficacy.

1. Nagel's Model

Up to this point, I have been discussing the notion of reduction in a fairly informal way. I would like to pause for a moment here to look more carefully at that notion and to make it clearer. If I am correct, many of the problems raised for the anti-reductionist by Kim and others are rooted in problems surrounding just what reduction entails. In what follows I will confine my discussion to the model of intertheoretic[67] reduction put forward by Ernest Nagel in his *The Structure of Science.*[68]

My reason for doing this is straightforward enough. I believe it is this model that has motivated the anti-reductionist component of functionalism. The main bone of

---

[67] Please note that in what follows the terms 'theoretical reduction' and 'intertheoretic reduction' will be used interchangeably.

[68] Nagel, Ernest. *The Structure of Science*, (Hackett: Indianapolis, 1979)

contention between the type identity theorist and the functionalist concerns the multiple realizability argument and its consequences for identifying mental types/kinds with physical types/kinds. But it is Nagel's model that brings into sharp relief the reasons why such a reduction is thought to be problematic on the functionalist view.

For the multiple realizability argument which is central to functionalism is thought to show that bridge laws between the mental and the physical are not supportable and hence, that the reduction of psychology to some physical science like neurobiology will not succeed. For Nagel, such bridge laws are essential to the reduction of psychology to some physical science. Hence, the irreducibility of mental properties conceived along functionalist lines is really an irreducibility of the predicates distinctive of psychology like "is a pain" to corresponding predicates in some physical vocabulary. It is then irreducibility in Nagel's sense that I take to be central to the anti-reductive component of functionalism.

As indicated in the introduction, reduction can take at least two forms: ontological or theoretical. Dualists, like Descartes, and critics of physicalism like Frank Jackson and

Thomas Nagel, claim that there can be no ontological reduction of the mental to the physical. What they are claiming is that such identifications are false.

Non-reductive physicalism, as it is understood here, does not involve resisting such ontological reductions--for this is precisely what token identifications between the mental and the physical amount to. And the antireductionism under discussion here is a version of token physicalism. Thus, it is primarily *theoretical* reduction that is at stake in this debate.

Ernest Nagel, in his *The Structure of Science*,[69] outlined a model of theoretical reduction that has become standard in debates over antireductionism. According to that model, the reduction of theories involves showing how a higher level theory T2 is really just a special case of a lower level theory T1. Showing how T2 can be deduced from T1 does this. For example, if psychology is reducible to neurobiology, psychological laws must be instances of neurobiological laws. For on the assumption that a theory is a body of laws, showing how one theory reduces to another is really just a matter of showing how the laws of the former

---

[69] *Ibid.*

can reduce to those of the latter. But this is itself just a matter of deducing the laws of T2 from those of T1.[70]

So suppose it is a law of psychology that when an organism is in pain (M1), it will engage in avoidance behavior (M2). We can symbolize this as follows:

(1) $M1 \rightarrow M2$

Our next step would be to show how this law is really just an instance of some law of neurobiology, our reducing theory (T1).

But in order to do this, we would need the terms of T2 and those of T1 to match, otherwise reduction would be impossible. Since what we want is a valid deduction we need a way of dealing with the T1 and T2 terms so that we can see how the conclusion is (at least implicitly) contained in the premises.[71] If our conclusion contains information not found

---

[70] Nagel writes: "A reduction is effected when the experimental laws of the secondary science are shown to be the logical consequences of the theoretical assumptions...of the primary science," on p.352 of *The Structure of Science*.

[71] Although it should be noted that what is sometimes called the rule of addition allows us to derive the conclusion P v Q from the premises P alone. Hence from the statement John shot a film, we can validly infer that John shot a film or John went to the dance. Nagel recognizes this in a footnote

in the premises then our reduction cannot be deductive. On the view that intertheoretic reduction is a process of logical deduction, the premises (laws of the reducer) must contain the conclusion (laws of the reduced), albeit if only implicitly.

But terms like "pain," "itch," and "tickle," which are distinctive of psychological discourse, are not to be found in neurobiological theory. When theoretical terms are not shared across the reducing and the reducer theories, this is what Nagel calls the "heterogenous" situation.[72] He thus introduces the notion of a "bridge law," which is meant to link the terms of heterogenous theories so that the proposed reduction can be achieved. Bridge laws "bridge" gaps between the vocabularies of different theories by showing how apparently different predicates like "is a light wave" and "is electromagnetic radiation," for example, may be co-extensive.

---

on p.353 and dismisses it as irrelevant to the issue at hand since P v ~Q is equally derivable from P alone. The arbitrariness of this deductive inference makes it unfit for deducing known experimental laws.

[72] When the theories involved share the same vocabulary, this is called "homogenous" reduction. Nagel cites as an example the reduction of celestial mechanics to Newtonian physics.

In heterogenous reduction, these laws will typically take the form of biconditionals that imply that a given T2 term is applicable just when a T1 term is. Of course, this does not mean that the terms mean the same thing in the sense that they have the same dictionary definitions, or intensions. If they did, scientists could leave their labs and conduct much of their "research" by simply consulting the dictionary.

Rather, such bridge laws are indications of the empirical discovery that wherever a given T2 predicate is applicable, so a corresponding T1 predicate is also applicable. In this way, 'water' may not be synonymous with 'H20' though science tells us that these predicates are applicable in exactly the same circumstances. So, let Wx = "x is a sample of water"; and Hx = "x is a sample of H20." Then, $(x)(Wx \Leftrightarrow Hx)$, which can be read as "for all x, if and only if x is a sample water then it is also a sample of H20. Hence, a bridge law between a theory containing the term 'water' and chemistry would tell us that 'water' is applicable if and only if, or when and only when, the predicate 'H20' is applicable.

Let us return to our postulated law of psychology. It was:

(1) $M1 \rightarrow M2$

Where M1 = being an organism in pain and M2 = engaging in avoidance behavior. But this is equivalent to the expression:

$(x)(Ox \rightarrow Ax)$

Here Ox = "x is an organism in pain," and Ax = "x engages in avoidance behavior."

If a reduction is to go through we need to show how terms like 'being in pain' and 'engaging in avoidance behavior', which are not terms of neurobiology, can be co-extensive with terms of the latter. This is what constructing a bridge law consists in. So, if we could do the following, then perhaps a reduction between psychology and neurobiology would be in the offing. Let Px = "x has a C-fiber firing in its body," and Rx = "x is in some physical state R." Then:

(1) $M1 \Leftrightarrow \mathbf{P1}$ or (1) $(x)(Ox \Leftrightarrow Px)$

(2) M2 ⇔ **P2** or (2) (x) (Ax ⇔ Rx)[73]

Hence:

(3) **P1** → **P2** or (3) (x) (Px → Rx)

The idea here is that (1) there is some predicate or term **P1** of T1 (neurobiology) that is co-extensive with M1. So, for example, perhaps the predicates "is a pain" and "is a C-fiber firing" are co-extensive. Whenever you apply one, you are entitled to apply the other. This would link part of the otherwise heterogenous vocabularies of our two theories, priming them for reduction. It would be a bridge law in Nagel's sense.

Next, we need to show (2) that the consequent of our psychological law is co-extensive with some term **P2** in neurobiological theory. This would be another necessary bridge law. Suppose that engaging in avoidance behavior is co-extensive with one and only one such physical predicate; let us call it "R" for the sake of convenience. This predicate may be complex, but it ought to be clearly specifiable in neurobiological terms such that it can also

---

[73] Of course, this bridge law may be unnecessary since behavior might be thought to be sufficiently physical. But let us suppose that we wanted to reduce behavior neurophysiologically, in that case we might want precisely

be made co-extensive with the relevant psychological predicate. In other words, there ought to be some predicate of physical theory that can stand in for "R." If there is not, our attempt to construct a bridge law will have failed.[74]

Once we have demonstrated these bridge law linkages the derivation of T2 from T1 is possible. For we can see how (1) is an instance of, or deducible from, (2) and (3). For the predicate "being in a state of pain" is co-extensive with "being in a state in which a C-fiber firing takes place." And if "engaging in avoidance behavior" is really co-extensive with "being in some neurobiological or physical state R," then the law that states that being in pain leads to engaging in avoidance behavior can be understood as being an instance of the more basic law (3). But (3) is defined in terms of C-fiber firings and some other specifiable neurobiological predicate. Our original psychological law has been shown to be an instance of a purely physical law.

2. The Limits of the Model

---

such a bridge law.

[74] It is precisely this requirement that the MR argument blocks according to Fodor.

It is important to note before proceeding that this model, as a model of *intertheoretic* reduction, has obvious limits when it comes to settling the many metaphysical puzzles and questions that beset the mind-body problem. For example, Nagel takes up the following possible objection:[75]

> …[A] contemporary writer maintains that psychology is demonstrably an autonomous discipline with respect to physics because "a headache is not an arrangement or rearrangement of particles in one's cranium," and "our sensation of violet is not a change in the optic nerve." Accordingly, though the mind is said to be connected "mysteriously" with the physical processes, "it cannot be reduced to those processes, nor can it be explained by the laws of those processes."

He goes on to argue that the main problem with such an objection is that it assumes that reduction requires the logical derivation of one *property* or set of *properties* from

---

[75] Nagel, *The Structure of Science*. P.364.

another *property* or set of *properties*. But this is misleading according to Nagel, for:

> [I]t suggests that the question of whether one science is reducible to another is to be settled by inspecting the 'properties' or alleged 'natures' of things rather than by investigating the logical consequences of certain explicitly formulated *theories*.[76]

And, moreover, "it is not *properties* but statements which can be deduced."[77]

The idea here is that it is not a legitimate objection to a proposed reduction of T2 to T1 to argue that the properties of T2 cannot be deduced from those of T1. For example, one might insist that mental properties cannot be reduced to physical properties of a certain kind because the nature of the former is essentially perspectival hence cannot be deduced, or must at least be emergent from, physical properties. But Nagel is quick to point out that his concern is simply not with reductions of this sort. The

---

[76] *Ibid.*, p.368

idea at issue here is that of the logical entailment of the *statements* of T2 from the *statements* of T1.

Strictly speaking, properties do not get deduced from one another. Why not? Because deduction is a logical relation and logical relations obtain among statements, *not properties*. (Of course statements about such properties may or may not be deducible from one another.)

So, suppose it can be shown that there are bridge laws between predicates like "is a pain" and physical predicates like "is a C-fiber firing." Furthermore, let us suppose that the psychological laws that involve pain can be shown to be instances of purely physical laws, in the manner outlined above. Let us oversimplify matters a bit and suppose this sort of thing could be done for all psychological laws. This would be enough for the purposes of Nagelian reduction. Further metaphysical puzzles about how a pain could *be* a physical event of some kind, or how phenomenal feels could be properties of certain purely physical or neurobiological systems, are not relevant to the question of reduction in Nagel's sense.

---

[77] *Ibid*. P.368.

Moreover, Nagel does not think that generally speaking, arguments about the inability to predict the presence of a property from the presence of some other properties or states should be a definitive reason to preclude a possible reduction between two theories anyway. Nor should it lead to an "ontological hiatus" between the objects of the theories involved.[78] For it may turn out that future refinements in one theory will provide the means necessary for derivations of statements about properties necessary for reduction which have been unavailable until the present moment. Modern chemistry may be reducible to physics, but surely the chemistry of 200 years ago was not so reducible.

Moreover, we could not predict *a priori* for example, that the molecular structure of H20 would give rise to a substance that dissolves sugar but not gold. Similarly, we should not be surprised if mental properties (or, better, statements about such mental properties) like phenomenal feels, can not be "predicted" or "deduced" in a similar

---

[78] *Ibid.*, see for example pp. 365 where, considering the possibility of being unable to reduce statistical thermodynamics to some mechanistic theory because of discrepancies between the vocabularies of the two theories, he writes: "...[T]he impossibility of the reduction...follows from purely formal considerations, and not from some alleged ontological hiatus between the mechanical and the

manner from neurobiological, or otherwise strictly physical, facts.

We should also note that since Nagel is not concerned with the reduction of properties, but reductions of statements about properties, and hence predicates, misgivings of the irreducibility of properties are not really germane to the question of Nagel style reduction. For unlike Kim, who insists on the causal efficacy of mental properties, Nagel is concerned with the reduction of statements about events and entities, and the predicates used to refer to them--*not properties*.

So this means that if there is a bridge law between the predicates that refer to mental phenomena and corresponding physical predicates, this would be enough to predict the presence of the conditions which will satisfy the predicates "is a pain" or "is a physical state P" for any system with the relevant mental and physical characteristics. And this would set the groundwork for a reduction of psychology to some physical science.

When one theory is reduced to another theory, the resources of scientific inquiry limit the explanations

---

thermodynamical."

thereby afforded. Science can tell us why, given knowledge of the laws that govern the physical world, H20 can dissolve sugar but not gold. It can provide an explanation in terms of the molecular structures of the substances involved and further physical laws. But it cannot tell us why it is just these laws which are in place, and why these laws should have just the consequences they in fact have.

This means that reduction in the sense of the term we are using here is not an ontological relation. For if the primary targets of reduction are *statements*, not the objects, events, properties etc. described by those statements, it is hard to see how this could have any robust ontological import. Of course, one could insist that bridge laws be conceived of as identity statements.[79] Perhaps considerations of ontological economy or concerns about the causal closure of the physical might motivate such a move. But strictly speaking, there is nothing in Nagel's model to restrict bridge laws in this way.[80] If pains simply were

---

[79] As Fodor points out in his "Special Sciences," in *The Philosophy of Science*, Boyd, Gasper, and Trout (eds.), p. 431.

[80] See Fodor, Jerry, "Special Sciences, or the Disunity of Science as a Working Hypothesis," in *Synthese* 28 (1974), pp. 97-115 reprinted in *The Philosophy of Science*, (ed.) by Richard Boyd, Philip Gasper, and J.D. Trout, (M.I.T. Press:

physical states/events of some kind, this would explain bridge laws linking predicates like "is a pain" and "is a C-fiber firing." But lawful co-variations between these pains and C-fiber firings across a metaphysical divide would also serve to establish bridge laws of this kind. Why is this so? Let us turn to this question in the next section.

3. Reduction as Ontologically Neutral

The debate as we are considering it here concerns the question of whether a Nagel style reduction between the mental and the physical is possible or plausible.[81] But such a reduction is a primarily logical accomplishment--one that shows how T1 can be explicated by the more encompassing T2. As mentioned above, this is perhaps brought out most forcefully when we consider that the biconditionals that are necessary for reduction would obtain in any ontology that

---

Cambridge, 1991), p.430 where he makes this point. Jaegwon Kim also points this out in his *Mind in a Physical World*, chapter IV.

[81] I say "plausible" because it is always open to the reductionist to claim that even if current physical theory cannot reduce *current* psychological theory, as a matter of empirical fact, it could turn out that at some point in the future these theories might relate in a way amenable to the reduction of the latter.

contained lawful co-variations between mental events and physical events.

Let Px= "x is in pain," Rx= "x is in some physical state R," Wx= "x winces," Qx= "x is in some physical state Q." So, given a Leibnizian pre-established harmony, psychological laws could be deduced from physical laws since the lawful connections between the two domains would support bridge laws between them and enable us to show that whenever we had a psychological law like (1) above, according to which:

(1) $M1 \rightarrow M2$ or (1) $(x)(Px \rightarrow Wx)$

Then we also had bridge laws like:

(2) $M1 \Leftrightarrow P1$ or (2) $(x)(Px \Leftrightarrow Rx)$

(3) $M2 \Leftrightarrow P2$ or (3) $(x)(Wx \Leftrightarrow Qx)$

Hence, our physical law:

(4) $P1 \rightarrow P2$ Or (4) (x) $(Rx \Leftrightarrow Qx)$

So here is what we get. Our original psychological law was: for all x, if x is in pain then x will wince. But our bridge laws show that this law is really just a special case of a more basic physical law. But our bridge laws involve crossing a metaphysical divide. Because x is in pain if and only if x is in some physical state R, *but x's pain is a state of x's ectoplasm--it is not itself a physical entity.* As long as pains and some corresponding physical state lawfully covary, as a Leibnizian might insist, this would be sufficient to establish the biconditional: (x) (Px $\Leftrightarrow$ Rx).

For if there is a "pre-established harmony" between the mental and the physical, there will be some physical predicate which is satisfied when and only when pains occur. But this is consistent with being a pain being completely non-physical in nature.

If entailments of certain kinds between and among statements are all that are required by the model of reduction at issue, then ontological gaps as wide as those found in a Leibnizian ontology should be of no consequence. Of course, the spirit behind Nagel's model is undoubtedly a hardheaded scientific physicalism. This is why he assumes that bridge laws will require coextensive predicates. But as

we have just seen, such an assumption is not necessarily true. Coextensiveness is theoretically superfluous given the requirements of reduction as set forth by Nagel.

We can see this more clearly perhaps with a non-mental example. Let us imagine a law according to which whenever cats meow they extend their claws. Let us suppose also that cats meow when and only when dogs bark. Let's go one step further and suppose that whenever dogs bark, they wag their tails. Let us symbolize these laws as follows. Let the domain of x = dogs and the domain of y = cats. Also, let Mx = "x meows," Ex = "x extends its claws," Bx = "x barks," and Wx = "x wags its tail."

Consider the following laws:

L1. $(y)(My \rightarrow Ey)$

L2. $(x)(Bx \rightarrow Wx)$

L1 is the feline law that says that all cats extend their claws when they meow. Our second law is the canine law that says that all dogs wag their tail when they bark. Given the following bridge laws, we can reduce any laws stated in

terms of meows and barks to ones stated in terms of claw extensions and tail waggings.

BL1. $(x)(y)(Mx \Leftrightarrow By)$

BL2. $(x)(y)(Ey \Leftrightarrow Wx)$

Hence, we can deduce that whenever cats extend their claws, so dogs wag their tails:

RL1. $(x)(y)\ (Ey \rightarrow Wx)$

Here our ontologies are clearly distinct. Dogs and cats are distinct entities. Nevertheless, if the sorts of laws we imagined above were to obtain, we should be able to affect a Nagel style reduction by simply deducing laws about claw extensions and tail waggings from those concerning barks and meows. That the resulting law crosses two distinct ontological domains would not make the reduction any less valid.

Nagel's model requires that there be lawful regularities to support the biconditionals typical of bridge laws; but there is no reason to suppose that such bridge

laws will also be identity statements. The logical relations articulated by Nagel's model do not make any such demand. As we can see from this example, a lawful correlation between facts concerning ontologically distinct entities can support the sort of logical derivation required by Nagel's model.

This shows that reductions, as logical relations among statements, do not restrict in anyway our ontology. Remember, for Nagel we do not reduce properties or objects to other properties or objects. Rather, we reduce *statements* about those phenomena to other statements. Consequently, any realization of Nagel's model is ontologically neutral in the sense that it may be consistent with physicalism, but also with certain kinds of dualism, like strict parallelism. This is so because it is *statements*, not the things picked out by those statements, which stand in a reductive (i.e., deductive) relation.

These last few points are important because they show that the issue of causation is independent of the issue of reduction, in the sense here described. The reason for this is simple: if causation is an extensional, ontological relation, that is to say, a relation between *entities* like properties, objects, and/or events, then relations between

statements which *describe* those properties, objects and events, would seem completely irrelevant to questions concerning the nature of that relation. These relations (among statements about properties, objects, and events that figure into causal relations) are of course significant, but for purely logical, not ontological reasons. Whatever entities do engage in causal relations, it seems certain that the statements that pick them out and form the basis of theories that describe them cannot affect in any way the efficacy of such causal relata.

If Jones killed Smith, it does not matter whether Jones is described as Smith's brother, the tallest guy in town, or the youngest son of a Texan Minister. Changing descriptions of him from "the murderer of Smith," to "Smith's brother," for instance, would not alter the causal powers Jones has with respect to Smith's death. Of course it is interesting for us to know that the murderer of Smith is identical with Smith's brother; but this is possible because these predicates are not synonymous. It is not because, somehow, these descriptions affect in some way the causal powers of

the individual who happens to be picked out by them. As Davidson has reminded us:[82]

> [W]e must distinguish between the causes of events and the features we hit upon for describing them, and hence between the further question whether the events are characterized in such a way that we can deduce, or otherwise infer, from laws or other causal lore, that the relation was causal.

Reduction allows us to see how one apparently autonomous theory is really explicable in terms of some other, perhaps more comprehensive theory. It allows us to see how there may be descriptions of events such that "inferences from laws or other lore" demonstrate their causal relations. It does not, however, place any restrictions on the causal efficacy of any of the objects, events, properties etc. that may or may not be causally related to one another.

---

[82] In his "Causal Relations," in *Essays in Actions and Events*, (Oxford University Press: New York, 1980), p. 155.

For example, if psychology could be reduced to neurobiology this would show us how psychological laws could be cast in neurobiological terms, which would be an interesting explanatory achievement. For perhaps now psychological explanations could be recast in physical terms. But what entities in the world get pushed and pulled in myriad causal ways would undoubtedly remain unaffected by such an explanatory success. Why? Because the issue of reduction, as it is understood here, concerns *theories*. And as we have now seen, showing one theory to be reducible to another would not suffice to show that there were causal relations between the entities picked out by such theories. For it could turn out that Leibniz was right and that the mental and the physical never interact causally but merely maintain the appearance of such interaction through strict lawful connections between mental events and physical events. This would be consistent with the reduction of psychology to some physical science; but would not be sufficient to guarantee any claims about the causal efficacy of the mental with respect to the physical.

The Causal Closure Principle, which is a guiding tenet of physicalism, does place restrictions on the causal

relation. For it says that no physical to non-physical or non-physical to physical causal relations take place. It denies causal interactions that involve breaching the space-time nexus.

So this requires that each mental entity be identical with a physical entity if causation can take place. But the further demand that such entities have descriptions that are isomorphic with predicates in some physical science, such that the laws, predicates, etc. of psychology are shown to be instances of physical laws, predicates, etc. is entirely misplaced.

Consequently, the question of the irreducibility of the mental in Nagel's sense would not seem to affect the causal issue either way. For Nagel style reduction is consistent with reducibility across a metaphysical gulf, in which case reduction would not be sufficient to guarantee the causal efficacy of the mental conceived along dualist lines. (If we accept the Causal Closure Principle.) And, as I hope it is now clear, reducibility in Nagel's sense is not *necessary* for establishing the causal efficacy of properties, objects, etc. whose descriptions cannot be reduced to physical theory in the sort of systematic way Nagel's model requires. For

reducibility in this sense demands only that *statements* about the entities of the reducer theory, in our case psychology, be derivable from *statements* of the reducing theory, neurobiology.

But this is, as I shall show in the next section, really a battle over bridge laws. If bridge laws are unavailable to link the terms of psychology with those of some lower level theory then the reducibility of psychological theory to the latter will not go through. But if there are no bridge laws, this simply means that the predicates distinctive of psychology cannot be shown to be co-extensive with any non-gerrymandered sets of physical predicates. Hence psychological terms, and thus psychological theory, will not be reducible if the following argument is correct.

But this will not touch the issue of causation since the problem of what properties may be efficacious in the physical world, if limited by the Causal Closure Principle, can be solved by identifying tropes of such properties with tropes of physical properties.

So, even if phenomenal feels are multiply realized as a class, and hence, predicates which pick them out do not

admit of bridge laws with some physical science predicates, this does not mean that each phenomenal feel is not identical with *some* physical property trope or other. For a mental property trope, or property instance, can be identical with a given physical property trope or instance even if the predicates used to pick them out as a class do not match up so that they support a Nagelian bridge law. Hence, even if the set of such physical properties is too gerrymandered for bridge laws, and reduction cannot go through, we can still maintain that such mental properties are causally efficacious because identical with various instances of physical properties.

Let us now turn to Fodor's argument for the irreducibility of mental properties and see how this claim can be made consistent with the demands of mental causation.

Chapter IV: Fodor, the Irreducibility of the Mental, and Kim's Causal Preemption Argument

We have already noted that any proposed Nagel-reduction between two heterogenous theories T2 and T1 requires bridge laws linking the terms of those theories. So if psychology is to be Nagel-reducible to physics or neurobiology, then we will have to link distinctively psychological terms with distinctively physical or neurobiological terms via such bridge laws.

But the existence of such bridge laws is exactly what is ruled out by the multiple realizability (herein "MR")

argument that motivates functionalism. For according to MR, mental predicates cannot find a faithful mirroring in physical theory. Thus, the kinds of psychology, even if they range over purely physical tokens, cannot be linked with kinds at the physical level.

In his "Special Sciences, or The Disunity of Science as a Working Hypothesis,"[83] Jerry Fodor endorses this view, arguing that reductionism, as it has been traditionally conceived by philosophers of science, is based on a faulty inference. Namely, that the generality of physics (that all events are physical events) entails the *reducibility* of special science theories.

One of the main claims put forward by Fodor in support of this view is that reductionism is just too strong as an empirical hypothesis. This is so, not because the generality of physics is implausible. Rather, it is because the taxonomies of the special sciences simply do not, as a matter of fact, match up with the taxonomy of physics in the

---

[83] See Fodor, Jerry, "Special Sciences, or the Disunity of Science as a Working Hypothesis," in *Synthese* 28 (1974), pp. 97-115 reprinted in *The Philosophy of Science*, Richard Boyd, Philip Gasper, and J.D. Trout, (eds.), (M.I.T. Press: Cambridge, 1991)

right sort of way. It is worth quoting Fodor at length here. As he puts it, with regard to the case at hand:

> If psychology is reducible to neurology, then for every psychological kind predicate there is a co-extensive neurological kind predicate, and the generalization which states this co-extension is a law...[t]here are no firm data for any but the grossest correspondences between types of psychological states and types of neurological states, and it is entirely possible that the nervous systems of higher organisms characteristically achieve a given psychological end by a wide variety of neurological means. It is also possible that given neurological structures subserve many different psychological functions at different times, depending upon the character of the activities in which the organism is engaged...The present point is that the reductionist program in psychology is not to be defended on ontological grounds. Even if (token) psychological events are (token) neurological

> events, it does not follow that the kind predicates of psychology are co-extensive with the kind predicates of any other discipline (including physics). That is, the assumption that every psychological event is a physical event does not guarantee that physics (or, a fortiori, any other discipline more general than physics) can provide an appropriate vocabulary for psychological theories.[84]

There are several important points raised by Fodor here. Most importantly, anti-reductionism does not have to be an ontological thesis. One may argue against reductionism on the ground that the kinds carved out by higher level sciences, like psychology, are ontologically distinct from those posited by physics. As noted earlier, this is exactly what a Cartesian, Jacksonian, or (Thomas) Nagelian might claim. But as Fodor's reasoning suggests, this is not the only antireductionist option.

One may argue that kinds and predicates of special sciences like psychology simply do not form lawful co-

---

[84] *Ibid*, p.433 (in Boyd, Gasper, and Trout).

extensions with the kinds and predicates picked out by physical theory. This may be the case even if there is no ontological split between the two domains. As we saw earlier, in a more mundane illustration, one can deny that there are lawful co-extensions between being a red object and being a round object while still insisting in general that every colored object must have a shape. To deny that shapes and colors have law-like connections between them does not commit one to anything like an ontological divide between the "color domain" and the "shape domain."

Moreover, as we have already seen in a general way, Nagel's model does not place any significant ontological restrictions on the relations between theories involved in reduction and the entities they pick out. Bridge laws *can* be identity statements linking the ontologies of two theories, but they don't have to be. Similarly, the failure of reducibility between two theories needn't show that there is an "ontological hiatus" between them. [85] Otherwise it would

---

[85] C.f. Nagel's *Structure of Science*, pp. 365, where, considering the example of reducing phenomenological thermodynamics to statistical mechanics, he writes: "…[T]he impossibility of the reduction without such special hypotheses [i.e., bridge laws] follows from purely formal considerations, and not from some alleged ontological hiatus between the mechanical and the thermodynamical."

follow that any irreducible theory or "special science" must be about entities which were at a metaphysical remove from the rest of physical reality. But this is surely an absurd consequence. Bridge laws may fail even when the ontologies of the theories are not split apart--this is the main lesson of Fodor's above quoted passage.

So, the events, objects, and entities picked out by predicates of higher level disciplines might all fall within the physical framework that is described by basic physical theory even if there are no lawful correlations between applications of the higher level predicates and particular sets (or types) of lower level predicates. This allows for an anti-reductionism about the *theories* of the special sciences while, at the same time, also permitting an ontological unity among the objects and events described by those theories. To see this more clearly, let us consider one of Fodor's own examples.

In the paper under discussion, he considers the property of being a monetary exchange.[86] Now, if economic laws mentioning monetary exchanges were reducible to physical laws, we would have to show how bridge laws between

---

[86] On pp. 433-4 in Boyd, Gasper, and Trout.

the two domains could be supported. But because there seems to be no limit, specifiable in physical terms, as to what might or might not count as a monetary exchange it is quite hard to imagine what such a bridge law would be like. For monetary exchanges involving such physically diverse things as paper dollar bills, wampum shells, metal coins, plastic credit cards, electronic blips on ATM screens and so on would have to be describable in terms that constituted predicates which figure into the laws of physical theory.

Even if we supposed that the number of all monetary exchanges that ever exist will be finite and, hence, that we could put together a list of all the physical predicates/properties involved in such a stretch of space-time, we would not have anything like a Nagelian bridge law. For, as indicated in the above paragraph, Nagelian bridge laws require that we capture *lawful* regularities when moving from the higher to lower levels. And there is little reason to think that such a wildly disjunctive set of physical properties will have significant theoretical, and hence lawful, connections among them.

In this sense, what *counts* as a clearly specifiable predicate of economics cannot be shown to be reducible to a

predicate of physics. But this is not because economics involves non-physical entities. It is simply because the terms of economics, the concepts used to classify phenomena as economic, will not derivable from strictly physical concepts, terms, and taxonomy.

1. Supervenience

The sort of inter-level dependency hinted at above is often discussed in terms of supervenience. As I will be using the term, supervenience is a relation whereby a property or set of properties M may be said to supervene on another set of properties P, just in case there are no differences in M without corresponding differences in P. Davidson utilizes this notion of supervenience and describes its role in his thinking as follows:[87]

> Although the position I describe denies that there are psychophysical laws it is consistent with the view that mental characteristics are in some sense dependent, or supervenient, on physical characteristics. Such supervenience might be taken

[87] Davidson, Donald, "Mental Events," p.214 in *Essays on*

> to mean that there cannot be two events alike in all physical respects but differing in some mental respects, or that an object cannot alter in some mental respect without altering in some physical respect. Dependence or supervenience of this kind does not entail reducibility through law or definition

To put it another way, once a subvening set of properties (P) is put in place, then all its supervening properties (M) will be fixed.

So, if the mental supervenes on the physical then any two systems that are physically indiscernible will also be psychologically indiscernible. But notice: the dependency here does not go the other way. Two individuals may be in the same psychological state, and yet, may differ physically. One may be in human pain, the other may be in Martian pain, for example. Here, the mental property in question (being a pain) supervenes on some set of physical properties because the latter are necessary for the former. But this dependency relation does not admit of bridge laws

---

Actions & Events, (Oxford: New York, 1980).

whereby the two property/predicate sets may be thought to be in lawful co-extension with one another.

We will not get such bridge laws because what makes something a pain in general, according to the functionalist, can be abstracted from the physical details of any one of its implementations. Hence, simply knowing details about the physical constitution of some heretofore unobserved physical system, without knowing about its overall functional structure would leave unanswered the question "does it experience pain?"

Just as in the case of color and shape, where knowing about one property provides no evidence concerning the other property, merely knowing that something has, let's say, C-fibers, would not be sufficient for determining whether it had a mental life in which pain played any kind of role. For it might be the case that Martians have C-fibers even if these do not play the role in their physiological economy played by C-fiber firings in ours.[88] Similarly, MR shows that knowing that something experiences pain cannot tell us

---

[88] C.f. David Lewis' "Mad Pain and Martian Pain," on this point, in *The Nature of Mind*, David Rosenthal, (ed.), (New York: Oxford University Press, 1991).

anything about what kinds of physical underpinnings that mental fact must have.

Let us return for a moment to the uncontroversial case of color and shape to better illustrate this last point. As earlier noted, there may be no law-like correlations between colors, like green, and shapes, like sphericity.

Suppose that everything that is colored must have some shape or other. So, it might be that there are many green things that are spherical. Yet, we know that any such regularities are accidental. For they do not support counterfactual claims. That is to say, the mere fact that something is spherical would not provide grounds for expecting it to be green, or any other particular color for that matter. Nor does being green warrant the inference that something will be spherical. Even if every colored object must have a shape, it does not follow that there must be mappings between types of color and types of shape. In this way, there would be no "bridge laws" linking properties or predicates of the shape domain with those of the color domain. Any such linkages would be purely accidental. [89]

---

[89] Of course one notable difference between the mental case

In the case of the mental, types like pain are likely to be identical with heterogeneous sets of physical tokens. For this reason, they will not allow for counterfactual claims concerning the kinds picked out by psychological discourse. So, if being a pain is multiply realized, then its kindhood would be lost were psychology limited to facts about just *human* pain.

If the mental supervenes on the physical then mental properties, while dependent on physical properties, would be irreducible to physical properties. As Fodor describes it, mental kinds must be distinct from physical kinds precisely because of the independence that mental predicates have from physical predicates and descriptions:

> The reason it is unlikely that every kind corresponds to a physical kind is just that (a) interesting generalizations (e.g. counterfactual supporting generalizations) can often be made

---

and that of color and shape is that the color-shape relation is not one of supervenience, whereas as the mental-physical relations is. But this is not important since the example is meant merely to show how sets of co-extensional predicates can lack lawful or otherwise systematic correlations with one another while still being true of objects in the physical domain. In both cases this happens.

> about events whose physical descriptions have nothing in common; (b) it is often the case that whether the physical descriptions of the events subsumed by such generalizations have anything in common is, in an obvious sense, entirely irrelevant to the truth of the generalizations, or to their interestingness, or to their degree of confirmation, or, indeed, to any of their epistemologically interesting properties; and (c) the special sciences are very much in the business of formulating generalizations of this kind.[90]

We can formulate generalizations about colored objects that tell us little or nothing about the shapes of such objects. For example, it may be the case that to a normally sighted viewer under standard lighting conditions, any red object looks more like an orange object than a green object insofar as it is colored. Or perhaps any object that is blue must be darker than any object that is yellow.

Notice: these truths, if they are truths, would not depend in any way on corresponding truths about the shapes

---

[90] In "Special Sciences," in Boyd, Trout, and Gasper, p.433.

of such objects, even if every colored object will have some shape or other. There is no tracking of color predicates by shape predicates and vice versa. Analogously, Fodor argues that we can make generalizations about psychological predicates in the absence of corresponding generalizations at the physical level.

Psychological predicates then do not require isomorphisms or co-extensions with the predicates of physical theory. But this is consistent with the claim that psychological phenomena are physical. For this is just like the color example which shows that colors can be "shaped" even if those shape predicates do not run together in a systematic way with color predicates, such that bridge laws between them would be possible.

2. Kim's Causal Preemption Argument

Jaegwon Kim has developed a seemingly devastating line of argument designed to show that this sort of view of the mental/physical relationship cannot account for mental causation. In fact his argument, if successful, would show that any properties[91] that were not identical with physical

[91] We should note here that Kim's discussion centers on

properties would be epiphenomenal with respect to the physical realm. This goes for any properties of the so-called "special sciences." But in what follows I will focus on the case of mental properties and their causal efficacy.

According to Kim, any realization of a mental phenomenon like pain will involve a system's having the sort of causal structure required of the functionalist construal of pain. In humans it might be C-fibers that execute this causal role; but in other creatures it might be some other physical mechanism that does this job. The details of this mechanism are not thought to be important on Kim's construal of functionalism, since what makes a given psychological state what it is can be abstracted from the details of any one its particular implementations. This is just what the MR argument entails.

The problem arises when we reflect on the following two facts. On the one hand, if functionalism is correct, then psychological properties will be irreducible to physical properties because of MR. But in every case in which a mental property is realized, it will be because there is

---

properties. Nagel, as we have already seen, does not think that properties are the proper target of intertheoretic reduction. Fodor also avoids talk of properties in favor of

some physical system or component of that system that can implement the causal structure characteristic of the psychological property in question.

But Kim notes that when such a realization occurs, it is the causal properties of the realizing mechanism that are responsible for the causal powers of the system. He states this in his Principle of Causal Inheritance, according to which:

> If mental property M is realized in a system at t in virtue of physical realization base P, the causal powers of this instance of M are identical with the causal powers of P.[92]

What I take Kim to be saying here is something like the following. If being a pain is an MR property, then it has its causal efficacy only in and through the physical properties that constitute its realization base on a particular occasion. So, even if my Martian friend and I are in pain at exactly the same time, and this pain causes us

---

"predicates."

[92] Kim, Jaegwon. "The Myth of Non Reductive Physicalism," p.240 in *Supervenience and Mind* (Cambridge University Press:

both to wince, it would be the *physical properties* underlying each of our psychological lives that did the causal work in question. My C-fibers got me to wince but his M-fibers got him to wince, and these are both physical phenomena. When a property is multiply realized, it has whatever causal powers it has by virtue of being *realized physically*.

What follows from this principle, according to Kim, is the causal preemption of the mental properties of a given system by its various physical properties. This follows if we accept two other claims: the causal closure principle and the idea that when psycho-physical causation takes place, it involves a mental property having causal efficacy. According to Kim:

> Now we are ready to derive some consequences from these assumptions. Suppose that a certain event, in virtue of its mental property, causes a physical event. The causal closure of the physical domain says that this physical event must also have a physical cause. We may assume that this

---

New York, 1993).

> physical event, in virtue of its physical property, causes the physical event. The following question arises: What is the relationship between these two causes, one mental and the other physical? Each is claimed to be a cause of the physical effect.[93]

The only possibilities here seem to be causal overdetermination or outright preemption of the mental by the physical, since the mental property in question is dependent for its causal efficacy on its subvening physical base. But this seems to leave the mental with no causal work to do.

To see this point better, let us return to the example given above. Being a pain is MR, so since it is C-fibers that get me to wince, and M-fibers that get my Martian buddy to wince, it cannot be the property of being a pain that did the causal work. For that would require the reduction of being a pain to either being a C-fiber firing or being an M-fiber firing, or the disjunction of these two properties. But this is exactly what MR rules out.

---

[93] *Ibid.*

But if the property being a pain is not identical with the property being a C-fiber firing, or being an M-fiber firing, and if the latter two would be expected to do the causal work here, there is nothing left for the mental property in question to do. Hence, functionalism with its MR properties, leads us down the garden path to epiphenomenalism, according to Kim.

Alternatively, Kim suggests that we could deny the causal closure principle, insisting that mental properties can have causal influence on physical properties to which they are irreducible. But this would put us right back in Descartes' camp with a metaphysically bifurcated causal nexus. If the anti-reductionist can have mental causation within a causally closed world, he must find away around Kim's causal closure/causal preemption trap. Thus the problem of "Quasation" that was discussed in chapter II has re-surfaced.

3. Classificatory Mental Properties vs. Mental Tropes

What supervenes on the physical are *mental types*. Pain *as a type*, supervenes on the physical, but particular pains do not so supervene. They simply *are* various physical

tokens, as are their qualitative properties, which are identical with physical tropes. Let me explain this further.

The irreducibility component of functionalism, as we have already seen, derives from the alleged fact of MR. But MR is a problem that involves relating *types* of mental phenomena to *types* of physical phenomena. There is no hint in any of these arguments that there is something non-physical about any mental *tokens*. Thus, what it is for anything to be a pain, while irreducible because functional in nature and hence MR, will be physical in the wider sense that it does not involve a breach of the space-time nexus. That is, it will consist in being some physical token or set of tokens in any of its instances in the actual world. Similarly for the painfulness of any given pain state—it will be identical with some physical trope.

Painfulness *as a type* may supervene on the physical. This means that any given pain state must be a physical state even though these physical facts will not add up to a neat isomorphism between the physical and psychological domains whereby the types or kinds distinctive of each could be linked by counterfactual supporting bridge laws or identities. But any instance of painfulness will be

identical with some physical trope. Mental tropes do not supervene on the physical because mental tropes just *are* particular instances of mentality. They are the realizations that make up the MR of the mental.

This move makes a huge difference to the line of argument put forward by Kim. For now the sense of competition underscored by his reasoning is stripped of its urgency. When I experience a pain that gets me to wince on a given occasion, it is not the property of being a pain that I expect to do the causal work. I expect this painfulness, here and now, to do it. Painfulness as such is out of the causal running.

But then the mental trope in question can have causal efficacy if the realization base of painfulness can. For such instances or realizations are identical with the physical tropes that would normally be expected to have a causal impact on the world. We can get around causal closure here, because we have not left the physical domain in tracking down causes. We can avoid reducibility because the types in question are still merely *supervenient* on their physical realization bases. Further, we can ensure that a given mental cause had an effect in virtue of one of its

mental properties if that mental property is a trope that happens to be identical with a physical trope. And, as we saw earlier in this essay, such identification is entailed by (antireductive) physicalism.

Kim's argument then focuses on mental property universals. For it is these properties that are rendered irreducible by the functionalist MR argument. But if these properties should not be expected to do any causal work, the preemption problem does not pose a serious threat to the functionalist.

Furthermore, even if a type identity between a property like being a pain and being a C-fiber firing were possible this would not show that being a pain was causally efficacious. For it wouldn't be painfulness as such that was causally relevant in any given causal interaction between the mental and the physical. It would have to be *instances* of painfulness that were so involved. For it is not painfulness in general, but a particular painfulness that gets me to wince when I am burned.

But then the question of whether the mental universal in question can be reduced to a predicate, kind, or universal property of physical theory becomes irrelevant to

the question of mental causation. For it seems that we ought to be interested only in the realizations or "instantiations" of these properties.

But contrary to Kim, this does not show the mental to be epiphenomenal--it shows only the universals used in classifying mental phenomena to be inefficacious. Unless we restrict our notion of mental properties to these universals, and we have now seen that there are good reasons not to make such restrictions, the problem dissolves.

As Kim himself suggests, we want to be able to account for claims like the following: "when I walk to the water fountain for a drink of water, my legs move as they do in part because of my desire for water and my belief that there is water to be had at the water fountain."[94] But this seems perfectly compatible with antireductive physicalism.

If my desire for water and my belief that there is water in the water fountain are physical events/states of my central nervous system, it would seem odd to deny them causal efficacy. Of course there may be independent reasons to deny them such efficacy, but the fact that they happen to

---

[94] Kim, Jaegwon. "The Myth of Non Reductive Materialism," p.281

fall under mental predicates, or are instances of universal mental properties should not be one of them.

Of course Kim's concern is with the causally relevant features of such states/events. As he would have it, the mental properties here must somehow compete with their physical realizers for causal potency. He suggests that the solution here is the type identity move, for if we identify the competitors, the competition disappears:

> At this point you might protest: Why all this beating around the bush? Why not say the mental cause and the physical cause are the same? Identification simplifies ontology and gets rid of unwanted puzzles. Consider say that in this glass there are two distinct substances, H20 and water; that is, consider saying that water and H20 co-occur everywhere as a matter of law but they are distinct substances nonetheless. This would invite a host of unwanted puzzles...By dropping a lighted match in the water I extinguish it. What caused it? Was it the water or the H20? The identification of the water with the H20 puts all

> these questions to rest in a single stroke: there is here one thing, not two. The identity solution can work similar magic in our present case: the pain is a neural state - here there is one cause, not two. The limb motion was caused by the pain, that is to say, a neural state. The unwanted puzzles vanish.

But Kim's reasoning is confused. The confusion is that in order to account for the causal efficacy of the mental, not only must token pain states be physical; but the property *being a pain*, a universal classificatory property, must also be physical. Now this is problematic for several reasons.

As I have already argued it seems odd to expect classificatory properties to play a causal role in the sorts of cases at issue. This is the point made by Frank Jackson and Ausonio Marras discussed in chapter II.[95] Why should the property of "counting as a pain" make *any* difference to the causal powers of the thing in question? Functional properties like being a pain are properties that denote

---

[95] See pages 76-83 of this thesis.

prior causal capacities. It would therefore be unreasonable to expect such classifications to do causal work.

To see this point better, let us suppose for the sake of argument along with Kim that the MR argument is right. Consistent with this is the possibility that all human pains are neural states of some kind, and, also, that pains in other creatures might be differently realized. This means that, as a cause, any given pain state in a human will also be a neural state. Now why should the classification of pains in general bear, in any way, on the causal efficacy of such a state? If the type identity thesis is right, then all pains are neural states of this sort. But why should facts about *other* pains, or about pains *as a class*, or about the *classifications* of the entity that does the causing, make any difference to the causal powers of any particular pain in any given causal interaction?

"Identifying" pains with neural states, in the way that Kim suggests, only eliminates competition if it is supposed that such classificatory properties ought to have causal efficacy in the first place. But as we have already seen, such identifications concern the inter-relation between physical *theory* and psychological *theory*. They do not add to

or take away from the causal mix unless we were to think that somehow, the same neural state that got my arm to move, depended in some way on being classified in a certain way. Remember: the difference between type and token physicalism rests on a disagreement on how classifications of mental phenomena relate to classifications of physical phenomena. *Ontologically*, these views are in agreement.

As the example of human pain and Martian pain shows, it is the actual mental tropes that we would expect to do the causal work in question--the multiply realizable property is just a classificatory predicate. Viewing mental properties as physical tropes eliminates the need to recruit universals like being a pain into the causal framework.

A given pain might be a given neural state. There is no puzzle to solve, no simplification required. For now, causal relations involve particular instances of mentality which are, at the same time, instances of the physical. There is no competition here between a universal like being a pain, and a particular instance of pain. But this is consistent with recognizing the force of MR--for what it is for anything to be a pain might preclude the isomorphisms between psychology and the physical sciences that the type

theorist requires.

So, the problem according to Kim arises from thinking of universals, like being a pain, as irreducibly mental. But as his passage also suggests, we can dissolve the problem of mental causation by identifying a given pain token with a neural state. So there is no further need to identify *all pains* in such a manner, as Kim seems to think. Hence, there is really no good motivation to think that a type identity move is necessary to account for mental causation. Such necessity would follow only if our concern were to get mental property universals like being a pain into the causal scene. But we can resist this type identification move. And as we have seen, such irreducibility should pose no serious threat to the antireductionist since such properties, if universal in nature, *could not, in any case, be expected to do the causal work with which Kim is here concerned.*

This is so because of the following dilemma that theorists like Kim, who demand the causal efficacy of universals, face. Universals are either spatio-temporal or they are not. If they are not spatio-temporal, then they cannot be recruited into the causal nexus, on *any version of physicalism.* For, whether one is anti-reductionistic or not,

the causal closure principle precludes the causal efficacy of entities that are outside space and time. Even on the type identity view, universal mental properties, whether they track the properties of physical theory in a systematic way or not, will still be outside the physical, causal realm.

Thus, the move to an identity solution suggested by Kim is really not going to get us anywhere, as long as we are stuck with universals as our target mental properties. The problem here is not with the mental-physical contrast, but with the universal-particular contrast. If the causal closure principle rules out the causal intrusion of entities that are not spatio-temporal then it is difficult to see how incorporating universals into the causal matrix would not be as problematic as incorporating Cartesian substances or vitalistic entelechies, and so on.

Of course, one might, following Bertrand Russell, claim that universals are encountered in experience all the time, making their existence independent of spatio temporal facts, but allowing them to have spatio temporal instances. This would seem to allow us to have our universalist cake and eat it too, allowing for the causal efficacy of universals after

all. In this case, universals are spatio temporal in some sense. But now we face the second horn of the dilemma.

If universals themselves are outside of space and time but their instances are not, then the reference to such universals is irrelevant to the issue of mental causation. For it would then be the *instances* of such properties, not the universals themselves, that would be doing the causal work. Whether or not these universals were linked through bridge laws with physical property universals would seem irrelevant since this would only concern the question of whether there were theoretical isomorphisms between psychology and the physical sciences. It would not tell us that such universals, because tightly linked or even "identical with" physical properties, were causally efficacious. But it should be clear by now that this is not something for which we should be hoping anyway. For it is mental property tropes, not mental property universals, that must be incorporated into the causal game.

For suppose that the type identity theorist is right and being a pain is just being a C-fiber firing. Still, it would never be pain in general or being a C-fiber firing in general that ever got anything to move, or be otherwise

causally influential. It would be particular pains, and hence, particular C-fiber firings that did this work. Thus, focusing on the reduction of such properties in order to account for mental causation seems a fruitless endeavor.

4. Conclusion

Let us review the argument thus far. We have seen that the anti-reductive component of functionalism centers on the irreducibility of the *classificatory predicates* or *universal properties* of psychological theory. The idea is that these are irreducibly mental because of the MR argument. But this is consistent with the demands of mental causation because it still allows the causal efficacy of mental property tropes by identifying them with physical property tropes. We can grant the MR argument, and the claim that psychological predicates or property universals cannot be reduced to those of the physical sciences, and hence maintain a healthy respect for our antireductionist intuitions. At the same time, we can account for mental causation by identifying mental tropes with physical tropes.

The preemption problem dissolves once we see that *mental types* or *mental property universals* supervene on the

physical. I say the problem dissolves because now there is no real competition to motivate the preemption problem in the first place.

Mental types, as universals, are not in competition with physical properties. Such universals are outside the causal realm because they are not spatio-temporal, period. This is a fact that cuts across the type-identity/token-identity distinction. In any particular instance of a causal interaction between the mental and the physical, the mental properties involved will be dated particulars, all of which will be identical with various physical tropes. Hence, such mental properties do not "compete" with themselves, and may not be preempted by subvening physical properties.

For, at the level of tokens and tropes, which is the proper level for causal relations, there is no supervenience relation. Supervenience is a dyadic relation and is precluded by identity. Of course the identity is not between mental property universals or classificatory predicates and those of the physical domain, for these mental properties still merely *supervene* on the physical. Pain, *as a type*, will supervene on various physical bases, but particular

instances of painfulness just are various physical tropes—as tropes they do not supervene. What could it mean to say that some instance of painfulness supervenes on a physical token with which it is identical?

Furthermore, the worry raised by Paul Noordhof that we discussed in chapter two now dissolves. If you'll recall, that worry was that even if a given mental trope is identical with a given physical trope, its being mental plays no causal role. But given that being mental is a classificatory designation we should not expect it to play any causal role. If causation is extensional, how we describe things ought not to be expected to play any causal roles. If an instance of painfulness is identical with an instance of some physical property, then the former has all the causal efficacy of the latter. The question of its counting as mental playing a causal role is a red herring that betrays a commitment to an intensionalist notion of causation.

Hence this move is physicalist, but non-reductionist. The identity here is between mental property *tropes* and physical property *tropes*. This eliminates the "competition" that is suggested by Kim's arguments and thereby allows us

to be physicalists without being *reductive* physicalists. This seems, in itself, a compelling argument for the claim that the properties of mental causation are tropes.

Chapter V: Davidson and His Critics

In this chapter I will take a close look at Davidson's Anomalous Monism (herein "AM") and the charge that this view leads to a kind of property epiphenomenalism. Many of Davidson's critics concede that AM does not lead to event epiphenomenalism,[96] so the real problem for the anomalous monist, if the critics are right, is an alleged dualism between mental and physical properties and the epiphenomenalism to which this is thought to give rise. The difference between these two types of epiphenomenalism should be clear enough by now. Event epiphenomenalism claims

---

[96] See for example J. Kim 1995; E. Sosa 1995; B. McLaughlin, 1995; F. Jackson, 1996; Block, 1995.

that mental *events* are causally impotent. If event epiphenomenalism is true, the mental can exert absolutely no causal influence on the physical. Property epiphenomenalism, on the other hand, is somewhat narrower in scope. It allows for the causal efficacy of mental events, but denies that their specifically mental properties play any causal role. So even if a given mental event has some causal efficacy, it is not *because it is mental* that it has this potency. It is epiphenomenalism of this latter variety that Davidson has been accused of by many of his critics.

1. The Anomalous Nature of the Mental

Before turning to the claim, made by numerous philosophers, that Davidson's AM cannot account for the causal efficacy of mental properties, I would first like to discuss Davidson's arguments for the anomalousness of the mental. If I am right, it is this aspect of his view which is at the root of so much of the criticism that has beset AM as an account of the mind-body relation. Jerry Fodor, for example, has made the issue of psychological laws central to his disagreement with Davidson's AM. He has argued that in

order to make intentional properties causally responsible there must be "intentional causal laws...contrary to the doctrine called "Anomalous Monism"."[97] But just why does Davidson deny the existence of such laws? And would the existence of psychological or psychophysical laws make much of a difference to the claim that the mental is causally efficacious with respect to the physical? Must there be such laws in order for mental properties to be incorporated into the causal nexus of physical interactions?

Davidson's arguments concerning the anomalousness of the mental are notoriously obscure.[98] I will not pretend to present an interpretation that illuminates every aspect of his thoughts here. Rather I hope merely to bring into relief some of the larger themes of Davidson's thinking about the distinction between intentional discourse and the nature of the physical sciences and the resultant conceptual autonomy of the mental that informs AM.

To begin with, Davidson thinks that there is some essential difference between the practice of ascribing

---

[97] Fodor, Jerry. (1989)

[98] For good attempts to make sense of them see Jaegwon Kim's "Psychophysical Laws," in *Supervenience and Mind;* and Simon Evnine's *Donald Davidson*

intentional states to a person and making predictions of the sort that are commonplace in the physical sciences. In the case of the former, my attributions are governed by normative and holistic constraints, whereas in the case of the latter no such constraints present themselves.[99] As Davidson has famously put it:[100]

> Any effort at increasing the accuracy and power of a theory of behavior forces us to bring more and more of the whole system of the agent's beliefs and motives directly into account. But in inferring this system from the evidence, we necessarily impose conditions of coherence, rationality and consistency. These conditions have no echo in physical theory, which is why we can look for no more than rough correlations between psychological and physical phenomena.

---

[99] Though as Kim points out in his "Psycho Physical Laws" (from Supervenience and Mind, 1993) holism may infect the physical as well as the mental for Davidson. Although Kim claims this, it is not clear that Davidson does hold this. It would seem odd for Davidson to maintain the distinction between the mental and the physical, in part, on the basis of the holism of the former if in fact he felt that both were holistic in some sense.

[100] "Psychology as Philosophy," in Essays on Actions and

> Just as we cannot intelligibly assign a length to any object unless a comprehensive theory holds of objects of that sort, we cannot intelligibly attribute any propositional attitude to an agent except within the framework of a viable theory of his beliefs, desires, intentions, and decisions. There is no assigning beliefs to a person one by one on the basis of his verbal behavior, his choices or other local signs no matter how plain and evident, for we make sense of particular beliefs only as they cohere with other beliefs, with intentions, hopes, fears, and expectations and the rest. It is not merely, as with the measurement of length, that each case tests a theory and depends upon it, but that the content of a propositional attitude depends upon its place within the pattern.

When I attribute to you the belief that snow is white on the basis of your uttering certain sounds, my attribution is governed by several factors. One is that I take you to be a being with a mental life and not a mere automaton. The very

---

Events, p.231

same sounds I hear coming from your mouth might also come from some inanimate source, like a tape recorder, to which I would normally resist attributing any mental states at all. Moreover, I must assign to you an understanding of the concepts 'snow' and 'white', and a desire to be expressing just the belief I assign to you and not some other. You may be attempting to deceive me or you might simply be idiosyncratic in your use of these terms in which case the content I would normally assign to you on the basis of these very sounds would get things wrong. How do I decide that you are expressing the belief that snow is white when you utter the words "Snow is white"?

I do so by making certain assumptions about you and your background beliefs, desires, and intentions. I decide that you are indeed a person with a mental life, that you mean by these words roughly what I would mean by them, and furthermore that you are being sincere. But notice, my evidence here must be subject to the very same interpretational constraints that limited my original attribution of propositional content to you. I might judge that you are being sincere because I cannot detect any possible motive for deception in the report of such an

innocuous belief. But of course, only rational people need good motives, and perhaps you are not rational. Or perhaps you do have some good reason to deceive me, or the desire to do so but, because of some inconsistency on your part, you might tell me what you believe anyway. Here again, determining background motives, intentions, beliefs and desires, only makes sense on the assumption that you are largely rational and consistent. Otherwise, I have no starting point from which to interpret your utterances. For all I know you mean the sky is blue every time you say "snow is white." Or perhaps your uttering these sounds is an indication of your desire to eat ham and eggs. (It is important to note that Davidson's argument does not depend on whether we in fact proceed in the way he describes. In fact we rarely do. But if pressed to make sense of someone's behavior, especially aberrant or idiosyncratic behavior, we might well proceed in just this manner.)

Thus I eliminate these outlandish possibilities only because I don't attribute to you a block of background desires, beliefs, and so on, that would make such an attribution plausible. If, for example, I had good reason to believe that you were speaking some code language, or

that you believed uttering trivial truths normally leads to receiving a nice big plate of ham and eggs, I might attribute to you such a belief. In the absence of such an odd network of background propositional attitudes, I would refrain from ascribing to you just these beliefs. But the beliefs that I ultimately do ascribe to you are ascribed according to the very same process.

Underlying Davidson's thinking here is a line of argument that is strikingly similar to one that contributed to the demise of logical behaviorism. For behaviorists, the meanings of mental terms were thought to be reducible to, i.e. definable in terms of, non-mental, behavioral terms. So Mike's wanting to walk the dog could be defined in terms of either Mike's overt dog-walk-wanting behavior, or his dispositions to engage in such behavior. Mike's wanting to walk the dog is nothing more than these behaviors and behavioral dispositions.

Aside from the obvious fact that such analyses miss the subjective component of much of our mental lives (being in pain is simply more than being disposed to cry out, or engage in various kinds of pain behavior) these accounts were thought to be hopelessly circular. Suppose, for

example, that we ask Mike whether he would rather take Fido out for a walk or let Sam do so. Here we might, parodying the behaviorist, claim that Mike's reporting that he would prefer to walk the dog would provide a clear cut behavioral criterion of his mental state of wanting to walk the dog.[101] But the sounds uttered by Mike can be construed as an expression of his mental states, and hence as a "reporting," only if we assume that Mike understands what we are asking him, what he is saying in response, that he is not systematically irrational or otherwise inconsistent, that he wants to tell us the truth, that this desire is not overridden by conflicting desires, that he does not have perverse beliefs about the social function of language, and so on. But in making these assumptions we have smuggled in large chunks of Mike's mental life and have failed to provide an analysis of his mental state(s) in purely behavioral terms.

The reason for this failure, if Davidson is correct, stems primarily from the normative and holistic nature of intentional state attributions. If I try to assign an

---

[101] Of course it seems much more plausible to suppose that this behavior is a consequence of his mental state, though I will assume this much for the sake of argument.

intentional state to an agent by simply considering some small sample of overt behavior, whether written, verbal, or otherwise, I can only make sense of it *as behavior* if I assume a large network of background beliefs, desires, intentions, etc. And when I then try to interpret that behavior as the report of a belief with a particular content, I face the same problem. I must move from witnessing a series of physical noises, marks, and movements to perceiving a person expressing the belief that torturing the innocent for fun is evil, for example. But how do I fix such an interpretation without first setting forth all sorts of assumptions about the person in question, i.e. that person's current mental life, background desires, beliefs, etc.?

Well suppose that one's mental life were simply identical with some set of physical facts, like the brain states of that person. Here we might hope to gain solid footing in our struggle against the translational indeterminacy that, according to Davidson's analysis, threatens to affect even the most mundane and apparently

straightforward cases of communication (e.g., Question: "How are you?"; Response: "Fine, thanks.").

Suppose, for example, that we could establish an identity between the mental state of believing that water is wet and some neural state, say N. If such an identity could be made, intentional state attribution might seem atomistic. Given enough neurophysiological knowledge, we might be able to simply read a person's mental states right off his brain, as it were.

But this cannot happen, for at least two reasons. In the first place, we would have to answer the question of how we arrived at our initial identification of the belief that water is wet with some neural state N. So far we have seen that fixing the propositional attitude of a subject involves taking into account a large pattern of interwoven beliefs, desires, and background assumptions about the rationality of the subject in question. If we can pinpoint that particular neural state as having the particular content in question then we must have been able to ascertain with certainty the truth of all the background assumptions brought into the translational endeavor. It is unlikely that this could ever be done without circularity, since whatever evidence we

might put forward for the claim that the belief in question had the content it is reported to have would only be compelling to the extent that it also made the agent more rational and coherent.

But we should only think that the agent is rational and coherent if we want to, for example, explain just why this belief and/or behavior, and not some other, is being attributed to the agent. Hence whatever evidence tends to support the interpretation requires support of its own and so cannot rely on the content of the propositional state in question as evidence.

Perhaps more importantly, when we do arrive at such an interpretation, we find also that it is difficult to see how a physical reduction of the state in question could be successful. For instance, if the agent in question is attributed the belief that water is wet, and the belief that water is H2O we might expect him to believe also that H2O is wet. But if Davidson is right this is not quite correct. We'd say that the agent *should* believe that H2O is wet. If the agent is rational, then there are certain normative consequences to take into account here. One of them is that when a belief is attributed to an agent, on the background

assumption of rationality, then it is also to be expected that the agent will believe all obvious logical consequences of this belief.

Now suppose, for the sake of argument, that we found some "identity" between the beliefs "water is wet" and "H2O is water" and some neural states, N and N* respectively. The rationality constraint that is essential to intentional state attribution now loses its place in our alleged psychophysical identification. For now we have to make sense of the normativity of belief attribution in neurophysiological terms, i.e. in terms of the empirical discovery of states N and N*. But how could we ever do this? It would seem to betray the empirical nature of neuroscience to bring into the enterprise such a priori normative constraints.[102] A person who believes that water is wet and that H2O is water *should, if rational,* believe also that H2O is wet. But this is an a priori discovery. What neural states an individual has is, by contrast, an empirical discovery. So how could such normative considerations be incorporated into a strictly empirical discipline like neurophysiology? The physical sciences tell us what is the

---

[102] Simon Evnine makes this point in his *Donald Davidson*

case, and make predictions about the future on the basis of supposed laws and experimental findings. But nowhere in the physical sciences do we find the kind of normativity requirement that is essential to the intentional, according to Davidson.[103]

It is in this sense that psychology and physics have disparate and divergent commitments for Davidson. It is because of the inherently normative character of intentional discourse that the mental has no "echo" in physical theory.

2. The Charge of Epiphenomenalism

(Stanford University Press: Stanford, 1991).

[103] We could, like Quine, and later Dennett, use this as a reason for denying the ultimate reality of the intentional. But Davidson uses this as a reason to claim that the mental is simply irreducible to the physical. Of course the holism of Davidson's view renders somewhat implausible the token identity thesis for which he argues. After all, if the content of an intentional state varies according to the demands of interpretation and the web of accepted background attitudes, it is hard to see how any such state will be token identical with a brain state. This is, I shall argue the ultimate stumbling block for AM vis-a-vis the problem of mental causation. After all, the reference to the brain state seems totally irrelevant to the attribution of the attitude. It is almost as if Davidson throws in the token identity only because he thinks that without it, there could be no mental causation. Of course if philosophers like Dennett are right, such causation is really only apparent--it is the result of viewing human beings from the intentional stance.

As remarked above, Davidson's AM has been the subject of intense criticism for some time now. In fact there has been a growing consensus that AM cannot account for the causal efficacy of the mental. As Kim puts it "the fact is, that under Anomalous Monism, the mental does no causal work."[104] Why does Kim think this? To see what is at issue let's consider what AM says about the relationship between the mental and the physical.

For Davidson, AM is the consequence of a certain conflict. On the one hand it seems that mental events interact causally with physical events. But Davidson holds that wherever there is causality there must be strict deterministic laws to back these relations (The Principle of the Nomological Character of Causality).[105] As far I can tell, Davidson gives absolutely no argument for this claim. He merely asserts it. But he claims also that there can be no psychological or psychophysical laws (The Anomalism of the Mental), for which Davidson does give arguments, as we

---

[104] Kim, J. "The Myth of Non-Reductive Materialism," in *Supervenience and Mind*, p.269.

[105] Davidson, "Mental Events," reprinted in *Essays on Actions and Events*: 208. (Oxford and New York: Oxford University Press, 1980).

saw above. Hence, the mental is anomalous with respect to the physical.

But this looks like a serious problem for anyone who thinks that the mental and the physical are causally linked, as Davidson plainly does. How might this problem be resolved? Well, if every mental event were in fact token identical with some physical event, then we could understand how mental causation obtains--there ought to be no special problem explaining how the mental interacts causally with the physical, since on this view the mental is just a subspecies of the physical. What is especially interesting about this position is that it does not force implausible type identities of the sort endorsed by good old-fashioned identity theories like that of J.J.C Smart.[106] So intentional states like "believing in Santa Claus" can be realized by one neural configuration in me, another in you, and even differently in the same individual at different times, on Davidson's view. This comports well with the idea that the brain has a certain plasticity that allows it to

---

[106] Smart, J.J.C. See his "Sensations and Brain Processes," *Philosophical Review* 68 (1958): 141-156.

compensate for damage in one area or part by switching functions to other areas or parts.

The critics, however, have claimed that AM is problematic because it implies that the mental properties of mental events cannot be causally efficacious. Sosa makes the point as follows:

> I extend my hand because of a certain neurological event. That event is my sudden desire to quench my thirst. Thus if my grasping is caused by that neurological event, it's my sudden desire that caused my grasping...Assuming the anomalism of the mental, though my extending my hand is, in a certain sense, caused by my sudden desire to quench my thirst, it is not caused by my desire qua desire, but only by my desire *qua neurological* event of a certain sort...The being a desire of my desire has no causal relevance to my extending my hand (if the mental is indeed anomalous): if the event that is in fact my desire had not been my desire but had remained a neurological event of a certain sort, then it

would have caused the extending of my hand just the same. [107]

Or consider Kim's dismissal of AM:

> What role does mentality play on Davidson's Anomalous Monism...?...None whatever. For Anomalous Monism entails this: *the very same network of causal relations would obtain in Davidson's world if you were to redistribute the mental properties over events any way you like; you would not disturb a single causal relation if you randomly and arbitrarily reassigned mental properties to events, or even removed mentality entirely from the world*...events are causes or effects only as they instantiate physical laws and this means that an event's mental properties make no causal difference.[108]

---

[107] Sosa, E. "Mind-Body Interaction and Supervenient Causation," in *Midwest Studies in Philosophy* 9 (1984).

[108] Kim, J. "The Myth of Non-Reductive Materialism," in *Supervenience and Mind*, p.269.

Each of these criticisms points to the same problem. On the one hand, the AM theorist insists that causality requires strict laws and that the mental interacts causally with the physical; but on the other hand, he claims also that there can be, in principle, no psycho-physical, or psychological laws. So Davidson's token identity solution comes to this: mental events interact causally with physical events because they have physical descriptions, or fall under physical kinds. But this invites the question: what role, if any, do the mental features of events play in causal interactions between the mental and the physical?

As Kim and Sosa both point out, it seems difficult to see how it is the mental features of events that have causal efficacy if, as Davidson seems to also think, causality involves strict, *physical* laws. This difficulty is especially acute for Davidson because he is famous for defending the thesis that reasons are causes.[109] If a given desire is token identical with a particular bit of neurophysiology, and if that neurophysiology is responsible

---

[109] See his "Actions, Reasons, and Causes," in *Essays on Actions and Events*: 3-19.

for some motor output, the fact that this neurological event is also a desire seems irrelevant. If reasons are really causes this result would seem intolerable.

At this point it is probably fair to ask just what mental properties are thought to be rendered epiphenomenal according to this criticism. As Sosa describes it, it is the fact that some particular bit of neurophysiology is a desire that requires incorporation into the causal picture. But it seems odd to think that *this* is the sort of property one would be worried about in trying to defend the thesis that reasons are causes. For this is a general classificatory property. Being a desire is a kind of *designation* that a given mental state has in virtue of having certain intentional features, for Davidson. But if those intentional features can be causally efficacious then why should we be troubled by complaints about the causal impotence of *being a desire*?

Davidson suggests a line of defense along these lines in his "Thinking Causes"[110] where he argues that falling under a given mental type is a product of how a given event

---

[110] In *Mental Causation*, 1995, pp. 3-18.

is characterized. But such characterizations could not affect the causal efficacy of such a state, on the reasonable assumption that the causal relation is extensional in nature.[111] So, he claims that requiring the mental features of events to play a causal role in interactions with the physical is tantamount to making the causal relation intensional. This is so because it would require an event not merely to be causally efficacious, but to be causally efficacious *as falling under a mental type*, or *qua mental* as it is often put. Thus he writes, in response to Kim's criticism:

> What Kim asks us to 'remember' is that 'on Anomalous Monism, events are causes only as they instantiate laws.' This is not anything I have claimed. I could not have claimed it since given my concept of events and of causality, it makes no sense to speak of an event being a cause 'as' anything at all. [My view]...is formulated on the assumption that events are non-abstract particulars, and that causal relations are

---

[111] Davidson argues for this in his "Causal Relations," in *Essays in Actions and Events*, (Oxford University Press: New

> extensional relations between such events. In his article, Kim does not dispute these two theses. But then there is no room for a concept of 'cause as' which would make causality a relation among three or four entities rather than between two. On the view of events assumed here, it makes no more sense to say event *c* caused event *e* as instantiating law *l* than it makes to say *a* weighs less than *b* as belonging to sort *s*...We can say, if we please that...events instantiate a law only as described in one way rather than another, but we cannot say that an event caused another only as described.[112]

When a given event is described as a belief, it fails to admit of incorporation into physical theory. As we saw earlier in this chapter, Davidson thinks that the concepts or descriptions found in an intentionalistic vocabulary find no "echo" in the conceptual apparatus of the physical

---

York, 1980), p. 155.

[112] Op. Cit., p.6

sciences. So, describing an event as a propositional attitude of some kind commits us to a vocabulary that is incommensurable with the vocabulary of the physical sciences.

The incommensurability here however is one between disparate vocabularies or descriptions of the very same class of phenomena. In being classified one way or another, the events under consideration are not altered. They do not move from a position within the causal nexus of the physical to an idle position outside it. Rather, one set of descriptions (the physical or neurophysiological) allows us to understand how a given event may fall under strict laws and hence, be engaged causally with other physical events. The other mode of characterization allows us to understand how attributing a given intentional state to an individual makes sense of his or her overall mental life, including the background desires, beliefs, and so on of the agent. But in moving from each of these descriptions, we are merely considering the very same event from different theoretical standpoints, on Davidson's view. The very same emerald ring can be considered qua colored, qua having such-and-such a weight, qua anniversary present, qua fifteenth object to

counted by Smith today, and so on. Similarly, a given event can be considered qua bodily event of some kind or qua intentional event of some kind for Davidson. But in either case, the causal powers of the event in question are unaltered by these varying descriptions. An emerald ring has whatever causal properties it has whether it is picked out as the most expensive item for sale at Tiffany's or Samantha's anniversary present, or etc. The case is precisely the same for mental and physical descriptions of events for Davidson.

In switching vocabularies then, an epistemological change is made: we can see how the event in question falls under deterministic laws, for example, when that event is picked out as a physical event of some sort, as opposed to a propositional attitude of some kind. But no metaphysical change is made here according to Davidson. The event classified is what it is, causal powers and all, regardless of this epistemological slide.

But Kim and Sosa argue that if the mental properties of various physical/mental events were stripped from the universe, not a single causal relation would be disturbed. This would seem to spell real trouble for Davidson's view.

As Sosa describes it:

> [I]f the event that is...my desire had not been my desire but had remained a neurological event of a certain sort, then it would have caused my extending just the same.[113]

This seems right, if we overlook the fact that the event that was my desire was also identical with the neurophysiological event in question. They are token identical for Davidson. Hence "taking away" the desire here would be equivalent to taking away the neurophysiological event. Of course, Davidson is not saying that there is a type identity between the mental and physical, but that does not mean that there is some metaphysical distinction between the mental event and the physical event tokens in question. Whether mental events reduce to physical events as types or only as tokens, it is still the case that necessarily, any given event is self-identical. Davidson's claim is simply that there can be no type reduction between the

---

[113] Ibid., p. 20.

classifications of psychology and those of neurophysiology. But he also claims that every psychological event will be a neurophysiological event.

Of course the point of these criticisms is that it is difficult to see just how there is any genuine behavior in the actual world given Davidson's view. Why think that the mental properties of events have any causal relevance whatsoever if the laws that back causal relations make no mention of such properties? We need a way of showing how it is these properties, and not others that are efficacious in the sub class of bodily motions that are actions. In response to Davidson's "Thinking Causes" Kim puts the point as follows:

> It must be admitted that Davidson's critics have not always been careful to distinguish between the following two claims: (1) *AM+P* entails the causal inertness of mental properties, and (2) *AM+P* fails to provide mental properties with a causal role. According to Davidson, (1) is false; and in this he is arguably right. However, this does not necessarily absolve *AM+P* of the charge of epiphenomenalism; for if something purports to be

a theory of mental causation assigns no causal role to mental properties--if it has nothing to say about the causal powers of mental properties while saying plenty about those of physical properties--the theory can, it seems to me, reasonably be said to be epiphenomenalistic with regard to mental properties. Plainly (2) is true, and has never been seriously disputed; and the defenders of *AM* have focused, by and large, on extending *AM* by adding a positive account of the causal efficacy of mental properties. In fact, that is Davidson's own approach in 'Thinking Causes': he wants to supplement *AM+P with supervenience (S)*, and perhaps also with 'non-strict laws', to restore causal efficacy of mental properties, tacitly acknowledging that within the framework of *AM+P* mental properties have no causal role to play.[114]

---

[114] Ibid., p.20

According to Kim then, Davidson is not guilty of event epiphenomenalism. But he is guilty of a certain property epiphenomenalism. After all, doesn't it make sense to ask of any given causal interaction, which of the properties involved are relevant and which are not? As Kim notes, there is an important difference between the causal efficacy of a pill's chemical properties and that of its color or shape. Doesn't it therefore make sense to ask just which of an event's properties are causally relevant and which are not? Can the anomalous monist account for the causal efficacy of mental properties? Or does his view simply underscore the causal priority of the physical properties of such events?

If Davidson is right, a given mental event cannot be incorporated into physical laws, as described. But then this classificatory mental property is irreducibly mental and hence outside the causal nexus. However, the intentional properties of the event must be identical with some component of the space time realm (and presumably some portion of the subject's brain). So the directedness of a given desire must be physical in nature, although when it is described as such, say as neural state N, it no longer

appears as a state *in which the subject should be* under the assumption of general rationality and the set of background beliefs and desires also attributed to him or her. Rather, this might be a state that can be predicted given some antecedent neurological conditions. But it will not be a state in which we think the subject must be in order to be largely rational, for instance. Whether the person is rational or not is not a question answered by neurophysiological investigations, if Davidson is correct. Hence, described as a physical state or event, the propositional attitude with which we might be concerned is no longer thought of as a state that adheres to the requirements of rationality and holism that define the nature of the intentional. But these changes in characterization cannot affect the causal powers of the event so classified, as we saw above.

The problem then is not incorporating the "mental qua mental," as Sosa complains, since the fact of being mental is a general classificatory property whose causal relevance seems negligible in defending the reasons as causes thesis. Remember, Sosa demands, not that Davidson accounts for how a desire can be causally efficacious in virtue of its semantic

or intentional features, but how "being a desire" can be a causally efficacious and relevant feature of the event that has this property. There is a subtle difference here. What Sosa is asking for is that the classificatory designation of a given mental event plays a causal role. What I am suggesting is that this is off the mark since what common sense really demands is that it is the intentionality of a given event that ought to play a causal role. It is not being a desire in general, but the intentionality of a given desire token that ought to play such a role.

What would really upset this view of mental causation would be to show that the same event which is my desiring this cup of water did not depend at all on its being directed toward the cup in a particular way. But if the token identity thesis is true, this property of directedness ought to be identical with some component of the neurological event. So it is unclear how one could take away the event with this directedness without also taking away the neurological event, since the two are identical on Davidson's view. Why then has Davidson's view been rejected by so many on the ground that it cannot account for the causal efficacy of mental properties?

I believe there are good reasons to be suspicious of Davidson here, and ultimately, it is the anomalousness of the mental that will spell disaster for his token identity solution to the problem of mental causation. Kim, Sosa, and company are right, but for the wrong reasons.

3. The Tension in Davidson's View

As we saw earlier in this chapter there is a tendency in Davidson's discussion of the anomalous nature of the mental to refer to the mental, as a conceptual category, in terms of the attributions of propositional attitudes that are made from the third personal perspective. As noted above, at least part of the reason for Davidson's denial of strict psychophysical laws has to do with the fact that the attribution of a mental state/event must meet certain normative and holistic constraints. What allows me to determine the contents of your mental states, for example, is an act of interpretation in which I try to make sense of your overt behavior by incorporating it into what I take to be your overall network of beliefs and desires. This of course must be supplemented by the (perhaps charitable) assumption that you are approximately rational. Here we

face a problem similar to that posed by Quine's radical translation scenario.[115]

For Quine, as well as Davidson, when we are confronted with the task of making sense of our fellow human beings, we must move from a paucity of observable empirical data, to a rich, highly determinate rendering of those data in mental/semantic/behavioral terms. Of course for Quine, such renderings are never fully determined by the evidence since in principle, conflicting translations that match perfectly the behavioral evidence can always arise. So when a native utters "Gavagai!" in the presence of a rabbit, I can interpret this as "rabbit," but the empirical evidence would also support equally translating this as "undetached rabbit part," "rabbity mass," "instantiation of rabbit-hood," and so on. Each of these interpretations can be made sense of given the totality of behavioral evidence.

In settling on any one interpretation, I assume a certain background ontology against which the attribution makes sense. But for Quine, such translation is not merely *underdetermined* by the available behavioral and otherwise observable evidence. There simply is *no fact of the matter*

---

[115] W.V.O. Quine, *Word and Object*, chapter two.

as to just what is meant by any particular utterance or other bit of linguistic behavior. For Quine, radical translation shows that there are no independent linguistic controls for a science of semantics. Consequently, Quine is eliminativistic about the mental, particularly the intentional.

Analogously to Quine, Davidson claims that whenever I try to make sense of the utterances of fellow human beings, I can do so only by assuming a large background--one of consistent propositional attitudes that render intelligible the ascription I in fact make in any particular case. What fixes my interpretation is determined by these constraining assumptions. But there is a problem here. For as I noted above, when we try to determine as a matter of fact just what a subject means by a particular utterance, new information about background propositional attitudes can undermine our original interpretation.

So I hear Philip utter the words "snow is white" and come to believe that he has expressed sincerely the belief that snow is white. But I then come to learn that Philip has also uttered to Samantha the words "snow is consumed by

humans, though it is produced by cows on farms." I now have reason to think that by uttering these words, Philip instead meant "milk is white." But this is not my only choice. I can infer that Philip thinks of snow as a liquid that is consumed by humans and produced by cows. In this case I would essentially be ascribing to Philip the belief that snow is milk, but this would make him irrational, or very ignorant, stupid, etc. There is a perhaps infinite number of interpretations that I can make on the basis of these two bits of data, but in the end I will choose the interpretation that makes the most sense, given whatever else I take Philip to believe, desire, etc.

Here we face a choice. The requirements of interpretation that inform the essential nature of intentional discourse for Davidson lead either to a Quinean irrealism about the mental or, a la Kantian noumena, to an epistemological block concerning the precise propositional contents of a subject's mental states.

But in either case it seems that it is from the third personal perspective that propositional content is either: 1)attributed to and therefore *determined by* the ascriptions of an interpreter; or 2) "out there" independently of this

ascription, but capable of being inferred only on the assumption of general rationality and certain background propositional attitudes. According to the first possibility, there would be no fact of the matter as to the contents of a subject's intentional states over and above the attributions made from the perspective of an interpreter.[116] According to the second option, we may not be able to prove that an agent is rational even though he is; but this assumption can help provide us with indirect knowledge of his intentional states.

Davidson's own discussion of these questions often obscures what is at issue. In defending the thesis that reasons are causes Davidson talks of mental states/events, and hence of propositional content, in realist terms: mental events are events in the physical causal chains that are involved in action. Presumably they are just brain states of persons. But in discussing the anomalousness of the mental Davidson's rhetoric has distinctly irrealist undertones and the focus on the third personal perspective should make one wonder just how the contents of one's

---

[116] This is essentially Dennett's move in *The Intentional Stance* (1987).

intentional states could be causally efficacious if they are subject to the vagaries of interpretation. After all, if what I mean when I say "snow is white" is *determined* by what an interpreter makes of the utterance, and if the possibility of conflicting, but equally plausible interpretations arises, it is difficult to see how content, if it is constrained by these third personal attributions, could be causally efficacious with respect to the central nervous system, the muscles, bones, and limbs of my body that are affected in action.

What Davidson actually thinks here is, I think, very difficult to decide. So I will not attempt to do so. In "Mental Events" he writes "...on the proposed test of the mental, the distinguishing feature of the mental is not that it is private, subjective, or immaterial, but that it exhibits what Brentano called intentionality."[117] Hence I think the defender of Davidson must insist that he is a realist about content, and furthermore, that there must be some determinate fact about what a given mental state is about.

---

[117] "Mental Events," in *Essays on Actions and Events*, 1980, p.211.

However, the holism of mental state attribution makes this token identity move very implausible, because the very content of an intentional state varies according to the demands of interpretation and the perhaps ever-changing web of accepted background attitudes, motives, etc. that are attributed *en bloc* to an agent on a given occasion of interpretation. Hence it is hard to see how any such state will be token identical with a brain state. After all, the reference to the brain state is totally irrelevant to the attribution of the attitude. Knowledge of a given agent's neurophysiology is neither necessary *nor sufficient* for knowing any of his or her intentional states. Not even a Laplacean super-neurophysiologist could tell me if I believe that the sky is blue by simply examining my brain states. He would have to engage in the same interpretive endeavor in which you and I engage whenever we attempt to communicate with others. Davidson throws in the token identity thesis only because he thinks that without it, there could be no mental causation. So unless a mental event is a physical event, it could not have a causal impact on the physical. But if I trace the causal ancestry of a bodily movement, like raising my arm, I will end up with a neurophysiological

event as my cause. This event would then have to be identified with whatever mental event was supposed to cause the raising of my arm. But the holism and normativity of the mental would seem totally out of place here, for we can find the cause without bringing to bear any of those considerations. Hence to pinpoint the neurophysiological event as the mental event here suggests either that such normative and holistic concerns are irrelevant or that they do not help in providing informative causal explanations. Either way, Davidson does not seem to provide a satisfactory account of the mental-physical relation, and thus, of mental causation.

Of course, Davidson would argue that to understand what sort of intentional state we were looking for in the first place, we would have to draw on the normative and holistic considerations that are essential to the mental. Moreover, he would claim that the cause here *just is* the neurophysiological entity, picked out under a different description.

But the anomalousness of the mental can give rise to conflicting intentional state attributions given the indeterminate and holistic nature of interpretation, as we

saw earlier in our discussion. Hence, it would be a real leap of faith to expect a neat mapping of such interpretational results onto the brain states that are thought to give rise to bodily motions in behavior. Where in the neurophysiological chain, for instance, does my belief that I can raise my arm end and my desire to raise it begin? Does my belief that I even have an arm exist somewhere in this causal chain as well? It is an obvious logical implication of my belief that I can raise my arm, and so the constraints of normativity and rationality would demand that I do hold this belief. Is it causally relevant here? If so, is it next to, before, or does it overlap with, my intending to raise a part of my body? Does my desire to raise my arm exist in every part of the neurological chain that triggers the movement of my arm? Or is it only in the part of that chain that exists in my brain? Presumably I have this desire at least up until the moment it is realized. How about the neurological transaction in the motor cells that get my muscles to contract? Do my beliefs and desires exist *there* as well? Do they "move along" the causal chain as one synaptic firing triggers another?

Of course these questions seem silly, given what Davidson argues about the anomalous nature of the mental. But it is this very anomalousness that makes his token identifications seem so strained.

I believe this fact underlies the widespread dissatisfaction with Davidson's solution to the problem of mental causation that has been expressed by so many philosophers in the literature addressing this topic.[118] Davidson does not fail to provide a workable account of mental causation in failing to demonstrate the causal efficacy of the mental qua mental. As we have seen, Davidson is correct in claiming that this requirement would involve making the causal relation intensional—a surely absurd consequence.

Rather, it is the emphasis he places on the holistic and normative nature of intentional state attribution that makes his token identity thesis seem a strained, almost *ad hoc* maneuver. It is this aspect of his thinking that

---

[118] C.f. F. Stoutland's "The Causation of Behavior," in *Essays on Wittgenstein in Honor of G.H. von Wright* (Acta Philosophica Fennica, 1976) p. 307; Dagfinn Follesdall, "Causation and Explanation: A Problem in Davidson's View on Action and Mind," in *Actions and Events: Perspectives on the Philosophy of Donald Davidson*; Mark Johnston, "Why Mind Matters," *ibid.*, p.423; See also J. Kim 1995; E. Sosa 1995;

ultimately undermines the account if what I have argued is correct.

But this problem also indicates a limitation on the trope approach that I have been advocating throughout this thesis. For if we accept Davidson's arguments for the anomalous nature of the mental then the trope move will not do us much good. For unless we can establish a plausible token identity between intentional states and neural states the requisite tropes will be unavailable. For we do not want mere classificatory tropes here. We would need particular instances of intentionality to be identical with particular instances of neural properties. But this seems wildly implausible given Davidson's arguments for holism.

And this challenge comes from another quarter as well. If, as we considered in chapter II, the content of my intentional states depends on factors outside my body, and if we have reason to think that the sources of movement in my body are internal to it, then intentional properties seem to fall by the causal way-side as well. For if my neural firings get my body to move as it does, but what is in my environment determines whether I have a desire for beer or a

---

B. McLaughlin, 1995; F. Jackson, 1996; Block, 1995.

desire for Tweer, then you will recall, we concluded that my desire cannot cause my movement.[119] The reason for this is simple. What gets my limbs to move are the neural firings of my central nervous system. But what determines the content of my intentional state are facts about the environment. In a Tweer filled environment, I have Tweer desires, not beer desires. Hence, the intentional component of a mental state does not seem to be in the causal running. The problem here is not one of the classificatory vs. the non-classificatory distinction.

Rather it concerns the fact that on the one hand it is the neurological states and properties in my body which seem to do the causal work that gets my body to move. At the same time, it is various external factors that seem to individuate the content of my mental states. Like Davidson's AM, it is what goes on outside me that at least partly determines the content of my intentional states. This fact poses, in a slightly different way, another serious challenge to the token identity thesis. For now, brain states do not seem "wide enough" to account for the content of my beliefs, desires, etc. But they do seem sufficient for

---

[119] See pages 47-48 of this thesis for the argument.

causing bodily movement. Hence it does not look as if content properties can be causally efficacious when we consider the bodily effects involved in movement. Whether there is a way around this problem is an interesting question.[120] But if what I have argued here is correct, Davidson does not provide it. Nor does the identification of mental tropes with physical tropes. Jackson and Pettit, along with LePore and Loewer offer some consolation with their various notions of causal relevance; but as we saw in chapter II, these solutions are consistent with epiphenomenalism.

The most a trope move can do, it seems, is to provide a place in the causal network for instances of phenomenal properties by identifying them with instances of physical, neural properties. Since these do not seem subject to the same sorts of normative, holistic, and externalist

---

[120] See in particular Jerry Fodor's *The Elm and The Expert* (1994) for a defense of intentional causation, though even he concedes that syntactic, and not semantic properties, do all the actual causal work. See also Stephen Stich's *From Folk Psychology to Cognitive Science* (1983) for an interesting discussion of the arguments in favor of a certain eliminativism concerning representational properties. For more along these lines, see also Daniel Dennett's *Intentional Stance* (1987). Frank Jackson and Philip Pettit also take up this issue, arguing that some content can be causally relevant in their "Some Content is

considerations which inform the attribution of intentional properties it seems these factors do not preclude a token identity of these states with physical states, thus securing their causal efficacy. Whether the intentional can be shown to be efficacious as well is a question beyond the scope of this thesis.

Narrow," in *Mental Causation* (1993).

Chapter VI: Mental Property Tropes and Subjectivity

> Supposing that there were a machine whose structure produced thought, sensation, and perception, we could conceive of it as increased in size with the same proportions until one was able to enter into its interior, as he would into a mill. Now, on going into it, he would find only pieces working upon one another, but never would he find anything to explain perception.
>
> --Leibniz, *Monadology*[121]

If Leibniz is right, there is something peculiar about the idea that minds might be identical with brains or any other "mechanical" phenomena. Even if every nook and cranny of my

---

[121] Leibniz, Gottfried Willhelm. *Monadology*, trans. G.R. Montgomery (La Salle, Illinois: Open Court Publishing Co., 1927) reprinted in *Eightennth Century Philosophy* ed. Lewis Beck (The Free Press: New York, 1966) p.196.

brain could be observed by someone, that third person knowledge would never include knowledge of the qualitative aspects of my experience.[122] Look all you like, you will never, on the basis of studying the physical properties of my brain, come to know what the mental properties of my experience are like. And if exhaustive knowledge of the physical properties of one's brain does not provide knowledge of the qualitative aspects of one's experience, how can such mental properties *be* physical properties?

My primary goal in this chapter is to take up the worry that just as intentional properties cannot be identified with physical properties of agents, so qualitative properties cannot be identified with properties of bodies. For if it is true that the qualitative features of itches, pains, tickles, perceptions, and so on cannot be identical with features of subjects' bodies, then trope theory will be unable to provide a solution to the mental causation problem. If mental property tropes of experience are not identical with physical property tropes then we will not have advanced much further than Cartesian dualism in trying

---

[122] Thus leaving what Joseph Levine calls the "explanatory gap" in his "Materialism and Qualia: The Explanatory Gap," in *Pacific Philosophical Quarterly*, 64 (1983).

to explain mind/body causal interaction. In order to solve this problem it must be shown that mental property instances *can* be identical with physical property instances.[123]

1. The Problem of Subjectivity

The arguments for the irreducibility of phenomenal mental properties are fairly well known by now. The two standard contemporary arguments given in this connection are put forward by Frank Jackson in his "What Mary Didn't Know"[124] and Thomas Nagel in his "What it is Like to be a Bat." [125]

In his ingenious thought experiment, Frank Jackson asks us to imagine a color scientist, Mary, who knows everything about the neurophysiology of color perception. She has a complete mastery of these neurophysiological phenomena. In fact, Jackson states as a premise in his argument that Mary knows *all* the relevant physical facts there are. As in Leibniz's thought experiment, we are asked to imagine that

---

[123] Of course whether they actually are is an empirical question.
[124] Jackson, Frank, "What Mary Didn't Know," in David Rosenthal (ed.), *The Nature of Mind*, (Oxford: New York, 1991), pp. 392-395.
[125] Nagel, Thomas, "What Is It Like to Be a Bat?" in David Rosenthal (ed.), *The Nature of Mind* (Oxford: New York,

every relevant piece of information accessible to the third person viewpoint is available.

However, Mary has been raised in a black and white laboratory and has never actually seen a colored object. When she is eventually let out of her achromatic environment, Mary comes to learn something new: what it is like to see colored things like tomatoes, grass, lemons, the sky, etc. She comes to have phenomenal knowledge where she had only physical knowledge. Up until the moment she is let out of her color-deprived world, Mary had unlimited knowledge of what goes on in people's brains when they perceive colored objects. But she learned nothing about what that experience was like in acquiring this neurophysiological knowledge.

This argument is supposed to show that mental properties cannot be identical to physical properties because even though Mary has exhaustive knowledge of physical properties, she nevertheless lacks knowledge of the qualitative aspects of color perception. Those qualitative aspects, the qualia involved in the perception of colored objects, resist a purely physical investigation. If in

---

1991), pp. 422-429.

knowing everything there is to know about the physical aspects of color perception Mary leaves out knowledge of what it is like to see colored objects, then, according to Jackson, such knowledge must involve non-physical properties. Look at the brain all you like, you will not see the perceptions of the subject under investigation, as Leibniz maintained over two hundred years ago.

Thomas Nagel, in his "What it is like to be a Bat"[126] makes a similar move. He argues that if even if we had exhaustive knowledge of the neurophysiology of a creature very different from ourselves, like a bat, our knowledge would be deficient. For we would never know, on the basis of this physical information, what it is like to be a bat. Hence, according to the present argument, the identification of the mental properties of experience with physical properties is impossible.

If Nagel, Jackson, and Leibniz are right, there is something peculiar about phenomenal qualities in that third person descriptions just cannot get at their essential nature. I cannot learn what it is like to taste a pineapple unless I actually bite into one. By contrast, I *can* learn

---

[126] *Ibid.*

everything there is to know about what physical changes take place in the body during taste experiences. But merely learning about the neurological changes in your brain when you eat a pineapple will never give me knowledge of the phenomenal character of that experience.

Even after I have tasted a pineapple, and have come to know what sorts of neurological changes take place in human brains as a result of tasting pineapples, I will not know *what it is like for you or anyone else to taste a pineapple.* Phenomenal mental qualities are not accessible from any standpoint but the first person perspective.

2. What These Arguments Show

The problem for the physicalist as we have outlined it is rooted in a certain epistemic asymmetry. The knowledge available from the first person viewpoint concerning the character of one's experiences seems inaccessible from the third person viewpoint. Why should this be the case? Throughout the physical world, we don't seem to encounter any such asymmetry.[127] Third person knowledge seems to

[127] Though, as Jim Landesman pointed out to me, we do seem to make some essential reference to consciousness, and thus to the first person perspective, in positioning ourselves in

exhaust what knowledge there is in routine physical investigations. Unless there is something it is like to be a piece of wood or a sample of gold, once we have mapped out all the physical features of the thing in question it does not seem like there is anything left out. Hence, we might be tempted to think that having an inner life is incompatible with physicalism.

But as Paul Churchland points out in his *Matter and Consciousness*,[128] this line of reasoning proves too much. For the difficulty concerning mental properties described by Jackson and Nagel remains even if we take a non-physicalistic perspective on the matter. Suppose, for instance, that we no longer think of mental properties as physical properties but rather as properties of Cartesian spirits.

---

time. That is to say, when someone tells me it is now 5:00 p.m. I can position myself in "the now" only by reference to my consciousness. Moments of time, in complete abstraction from consciousness, seem totally "now-less." Imagine a universe devoid of all sentience. What sense would it make to call any given moment of that universe's timeline "now?"

[128] Churchland, Paul, *Matter and Consciousness* (MIT Press: Cambridge, 1984). Churchland's positive thesis, however, is eliminative materialism. So he ultimately denies the existence of qualia. One virtue of the trope solution, as we shall see in the last section of this chapter, is that it does not require the denial of qualitative mental properties.

So, now Mary is a color scientist who comes to know everything there is to know about the changes in Cartesian ectoplasm that take place whenever bodies are placed in the vicinity of colored objects such that sensations of color arise in their connected, though non-spatial, minds. Notice: Mary will be as ignorant here as she was in our first scenario with respect to what it is like to perceive a colored object since her knowledge is limited to the experience of others. She cannot come to understand what it is like to experience the visual sensation of a red object, for instance, because her access is still limited to the third person perspective. She can come to know everything there is to know about the workings of Cartesian ectoplasm and still not know what it is like to perceive colored objects unless she does so herself. But the fact of whether her knowledge is of physical stuff or not becomes entirely irrelevant to the asymmetry problem.

And we can run this analysis *mutatis mutandis* for Nagel and his discussion of bat phenomenology. Imagine that we learn that bat qualia are identical with non-physical stuff.[129] Again, learning everything there is to know about

---

[129] Of course I have no idea how we would learn this, but

that stuff will not provide insight into the phenomenology thought to inform the mental lives of bats, and so on. Bat phenomenology would be forever beyond our grasp whether it was physicalistic or non-physicalistic in nature.

The foregoing, if correct, shows that the epistemic problem noted by Leibniz, Jackson, and Nagel, cannot be taken to have the metaphysical consequence that physicalism is false. For it would seem that first person/third person asymmetry is independent of the metaphysics of the situation. At the very least, it is neutral between physicalism and dualism. Against a dualist backdrop, the very same problem rears its ugly head. Cartesian spirits are at as much of a loss in trying to comprehend the experiences of other subjects as are fully physical creatures. But if this reasoning shows that there is an asymmetry between the first and third person perspectives whether or not we assume physicalism, then physicalism cannot be the culprit here. But if the problem is not physicalism, then what is it?

The problem is generated in thinking of points of view, and mental properties in general, in the way that universalists have traditionally thought of properties: as

---

let's assume this much is possible in entertaining the

real, unified entities that can somehow be *wholly* present across many instances at the same time. I do not experience pain-as-such, and so can never experience your pain. Your pain is one thing, mine is another, and never the 'twain shall meet. If the universalist is correct, then qualia should be *universal* properties, i.e. wholly present across different instances simultaneously. But this seems to conflict with the subjectivity of the mental as we have outlined it.

For if mental properties are inherently subjective then they should not be accessible from any viewpoint but the first person point of view. But if such properties were universal in nature, then they could be accessible from more than one point of view. A universal property is, by its very nature, repeated across numerous instances. For example, the property of redness, if construed along universalist lines, is present wherever a red thing is present. But it is the *same property* that is present across these numerically distinct instances. A red ball and a red car then have *the same* property of redness. If we now shift to mental

---

possibility that non-physicalism about qualia is correct.

properties, we can see how such universalism would conflict with the alleged subjectivity of the mental.

Suppose, for example, that the painfulness I experience at a given moment were a universal. If this were so, then we should expect that it could be present across numerous instances. But if the mental is inherently subjective, this result would be problematic; for it would allow for *my painfulness* to be present to more than one point of view. For just as redness is present wherever there are red things, so painfulness should be present wherever there are pain sensations. But clearly, if we accept the asymmetry of the third and first person viewpoints, my painfulness cannot be present to any point of view but my own. If that property were a universal however, it should be present whenever a pain sensation occurred--whether that pain was mine or someone else's. But then, if painfulness is truly universal there should be no epistemic asymmetry between the first and third person viewpoints. If we think there is such an asymmetry, the universalist theory will not do.

Of course one could argue that if we were to take, for example, Jones' pain, we could make out the case for universalism. We could do so by arguing simply that the

property of *being Jones' pain* is a property that recurs within the mental life of Jones. Hence, the property is both *universal* in being repeatable over the course of Jones' mental life, and *subjective* in being Jones' pain. Hence, there really is no conflict between the universality of mental properties and their subjectivity, as long one understands just how to individuate them. Universals may be in many places at one time, but they do not have be.[130]

While this response has a certain appeal, it is nonetheless problematic. For one thing, it leaves unanswered questions concerning general mental properties like being the awful feeling quale of pain. After all, why should we not be concerned with such general properties? Why limit our case to Jones' pain? If we are going to admit universals into out ontology, why be picky in this case? Limiting our investigation in this way indicates that there is something odd about thinking of mental properties in universalist terms. For on the universalist theory there will be some universal entity that is the feel of pain that can be repeated across myriad instances. But this seems in conflict

---

[130] For example, the Earth has only one moon, so the property of being the Earth's moon could be universal and have one and only one instance.

with the subjectivity of the mental. The move to restrict our property to "Jones' pain" looks like an attempt to sneak in a trope under the guise of universalism.

Of course, one might object that there is no real conflict here, since, for all we know, there is no problem with the notion that the feel of pain is repeated whenever someone experiences pain. Certainly, we may not be able to verify the qualitative similarity that these properties have across different individuals, but that would not show that such similarity does not exist. To think this would amount to a pernicious verificationism about the mental.

Furthermore, one might insist that it would not be the universals themselves that distinguish the qualitative aspects of different subjects' mental lives, but the *instances* of such universals. Since universals are instantiated in myriad instances, this would seem to make the universalist position perfectly capable of doing the explanatory work being demanded of it here. After all, Jones does not experience pain as such, but particular instances of pain. Hence it is unreasonable to expect universals to account for the first-person/third person asymmetry problem, for the universals are not what an individual experiences.

It is the instances of those universals that he does so experience.

The problem with this move is that it renders reference to the universal irrelevant. It would make such reference irrelevant, because by making the move we are essentially acknowledging that *instances* of universals, not universals themselves, play an active role in our mental lives. In that case we might as well dismiss the universals altogether for the purposes of trying to understand the nature of the mental properties that inform experience. If they play no role *qua* universals, why should we worry about them at all?

3. The Trope Solution

According to the trope analysis of mental properties, things like "what it is like" to taste a pineapple will be real aspects of the world, but they will be as particular as the objects they inform. Such properties should not therefore be expected to be accessible from the third person perspective because one man's pain is a numerically distinct entity from any other's on this view. Hence, no two people can experience the same mental properties. The reason for this is that, as tropes, mental properties will be totally

particular. Therefore, unlike universals, two qualitatively similar tropes will always be numerically distinct. For the universalist, similarity across instances is explained by the presence of a universal. But, as we saw above, this amounts to the claim that apparently distinct instances of a property are really instances of the *very same property*. Hence, even though a red car and a red ball are distinct objects, insofar as they are both red, they have the very same property of redness. On the trope view, two such similar objects have would be said to have qualitatively similar, though numerically distinct color properties.

So, in the case of mental properties like pain, I may be able to know what your pain is like by having the same sort of neurological changes that take place in your brain duplicated within my own. Of course this would not give me *your pain*, nor would it give me the very same *painfulness* that is tied to your point of view. My painfulness trope may be qualitatively similar to your painfulness trope, yet they will still be numerically distinct properties on the trope view.

The same would be true on a Cartesian view of the matter. If my painfulness is a matter of changes in my

ectoplasm, then you will never be able to feel my painfulness by simply coming to learn third personal facts about it. Part of the problem here is the logical point that once I experienced *your pain* it would cease to be *your pain;*[131] for it would now be a sensation of mine and therefore it would be *my pain*. It would therefore be as impossible for you to experience my pain as it is for you to win my race or catch my catches.[132]

But there is also another point to consider here. Let us suppose for a moment that phenomenal mental qualities are simply properties of one's central nervous system. In that case, there would be no more reason to believe that I can experience your qualia by observing them than there would be to think that I might be burned by watching a house go down in flames, or get fit by watching athletic events, or learn how to ride a bike by simply reading about it. Why not? Because the changes in my central nervous with which we are

---

[131] Of course there is always the logical possibility that changes in my brain or ectoplasm could give rise to qualitative mental states *in you*. I do not wish to deny this. But even if such a thing happened, it would be *your experience* of such changes that informed your consciousness. You could not, logically speaking, experience my experience or anyone else's.

[132] As Gilbert Ryle argues in *The Concept of Mind* (Harper & Row Publishers: New York, 1949) pp. 208-210.

identifying mental states and properties for the moment should not be expected to be brought about by my observing the effects of similar changes in another body, or reading about such changes in a book. When it comes to the physical changes in my brain that the physicalist equates with experience, I must have those changes take place *in* me (i.e., in my body) in order to have the experience in question. Reading about them, as Mary the color scientist might, should not be expected to give me the experience of color anymore than reading a menu should be expected to fill my belly. The propositional or observational knowledge I acquire from the third person standpoint should not be expected to affect me in the same way that having those changes in my body would. Reading about smoking does not give me cancer. Smoking does.

But this would not imply that such changes involve something non-physical. Your experience may very well be a result of changes in your central nervous system. It is also possible that they are completely non-physical occurrences--nothing in the arguments we have been considering so far rules this out. As we have noted, the epistemic problem

outlined above does not seem to have any clear implications for the question of physicalism versus dualism.

But still, a skeptical reader will ask: Are my experiences like yours? Could I ever, if mental properties were physical tropes answer this question? Moreover, why is there something it is like to be a creature with a mental life? Why should there be an "inner life" connected with having a central nervous system?

Here it must be conceded to the skeptic that the view I am defending cannot answer these questions. The trope view cannot explain why there is anything it is like to have experience. Nor can it explain the difference between non-mental properties, which have no trace of subjectivity, and for example, tropes of ticklish-ness, which seem to essentially.

But this does not seem a special problem for the physicalist claim that such properties are *physical properties*. For even if they were non-physical I would be equally ignorant as to whether my qualia resembled yours, why these qualia existed in the first place, and why things like specks of dust and rainstorms have no interior lives at

all. These questions remain, regardless of our choice of a metaphysic of mind.

I can never experience your pain on the trope view, not because there is something essentially non-physical about that experience but, rather, because like any other particular thing, your painfulness cannot be repeated across various instances. If the mental properties of experience are tropes then they will not be repeatable in the way that universals are thought to be. Like the instances of the universalist's ontology, each trope is a numerically distinct entity. Different tropes may be qualitatively similar, as we saw earlier; but they will nevertheless retain their numerical distinctness. If the properties of experience were universals then they could be wholly present across many different instances--in my brain/mind as well as any other. But if one man's pain is essentially tied to his point of view then this just cannot be; for the points of view which are thought to be essential to such experiences and the properties that inform them cannot be so well traveled.

That is to say, the painfulness I experience on a given occasion is a numerically distinct feature of my mental

life. You may not experience that self-same property. Though, it is perfectly conceivable that you might experience something qualitatively indistinguishable from it. This would be logically possible on the theory that the mental properties of experience are tropes. Though, again, trope theory cannot definitively decide whether such a resemblance obtains.

So the trope move does not explain why the world has subjective features.[133] That is, being a trope, does not in itself make something amenable to subjective access. Presumably, there is nothing it is like to be a weight trope, or a trope of yellow, for example. But the subjectivity of qualitative mental properties seems to require that they be local and particular. This is so whether we accept a thoroughgoing physicalism, partial physicalism, or full-blown dualism. One consequence of this view is that minds, whether conceived of physicalistically or not, are special in that they are the only things that can be both bearers and observers of the self-same trope. The trope view then is the *beginning* of wisdom, not an

---

[133] Thus, it does not fill Levine's "explanatory gap."

explanation of why or how such peculiar things as experience should exist.

Bibliography

Antony, Louise M. and Joseph Levine. "Reduction with Autonomy," in *Philosophical Perspectives*, 1997.

Armstrong, David M. "Against 'Ostrich' Nominalism: A Reply to Michael Devitt," in *Properties*, D.H. Mellor and Alex Oliver (eds.)

_____. "Properties," in *Properties*, D.H. Mellor and Alex Oliver (eds.) pp. 160-173.

_____. *What is a Law of Nature?* (Cambridge University Press: Cambridge, 1983).

_____. *Universals and Scientific Realism*, Vol.I, (Cambridge University Press: Cambridge, 1978).

_____. *Universals: An Opinionated Introduction*, (Westview Press: Boulder, 1989).

Baker, Lynne Rudder. "Metaphysics and Mental Causation," in *Mental Causation*, ed. Heil and Mele, pp. 75-97.

Beck, Lewis. ed. *Eighteenth Century Philosophy* (The Free Press: New York, 1966)

Blackburn, Simon. "Losing Your Mind: Physics, Identity, and Folk Burglar Prevention," in *The Future of Folk Psychology*, edited By John D. Greenwood, 1991.

Block, Ned. "Anti-Reductionism Strikes Back," in *Philosophical Perspectives*, 1997.

Burge, Tyler. "Mind-Body Causation and Explanatory Practice," in *Mental Causation*, ed. Heil and Mele, pp. 97-121

_____."Individualism and the Mental," in P. French, T Uehling, and H. Wettstein, eds. *Midwest Studies in Philosophy 4, Studies in Epistemology* (University of Minnesota Press: Minneapolis, 1979)

Campbell, Keith. *Abstract Particulars*, (Blackwell: New York, 1990)

Churchland, Paul. *Matter and Consciousness* (MIT Press: Cambridge, 1988).

Davidson, Donald. "Mental Events," first published in 1970 and reprinted in Davidson, Essays *on Actions and Events*.

_____. *Essays on Actions and Events* (Oxford University Press: Oxford and New York 1980).

_____. "Actions, Reasons, and Causes", *Journal of Philosophy* 60 (1963). Reprinted in *Essays on Actions and Events*.

_____. "Psychology as Philosophy," reprinted in *Essays on Actions and Events*.

_____. "Thinking Causes," in *Mental Causation*, ed. John Heil and Alfred Mele, 1995

Descartes, Rene. Meditations, in The *Philosophical Writings of Descartes*, vol. 2, ed. John Cottingham, Robert Stoothoff, and Dugald Murdoch (Cambridge University Press: Cambridge, 1985).

Dennett, Daniel C. *The Intentional Stance*. (MIT Press: Cambridge and London, 1989).

Devitt, Michael. "'Ostrich Nominalism' or 'Mirage Realism'?" in *Properties*, D.H. Mellor and Alex Oliver (eds.), pp.93-101.

Dretske, Fred. *Explaining Behavior*. (MIT Press: Cambridge, 1988).

Ehring, Douglas. "Mental Causation, Determinables, and Property Instances," in *Nous*, 30, 1996, pp.261-80.

Fodor, Jerry A. *The Elm and the Expert* (MIT Press: Cambridge 1994).

_____. "Making Mind Matter More," *Philosophical Topics* 17 (1989): 559-79; reprinted in *A Theory of Content and Other Essays* (MIT Press: Cambridge, 1990).

_____. "Special Sciences, or the Disunity of Science as a Working Hypothesis," *Synthese* 28 (1974): 97-115, reprinted in *Representations*.

Greenwood, John D., ed. *The Future of Folk Psychology* (Cambridge University Press: Cambridge, 1991).

Heil, John and Alfred Mele, eds. *Mental Causation* (Clarendon: Oxford, 1993).

Honderich, Ted. "The Argument for Anomalous Monism, *Analysis* 42 (1982): 59-64.

Horgan, Terence. "Kim on Mental Causation and Causal Exclusion," *Philosophical Perspectives*, 1997.

Jackson, Frank and Phillip Pettit. "Functionalism and Broad Content," *Mind* 97 (1988): 381-400.

"Some Content is Narrow," in *Mental Causation*.

Kim, Jaegwon. "Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism?" in *Mental Causation*

_____. "Concepts of Supervenience," reprinted in *Supervenience and Mind*.

_____. "Epiphenomenal and Supervenient Causation," reprinted in *Supervenience and Mind*.

_____. "Making Sense of Emergence," presented at the 1997 Oberlin Colloquim in Philosophy.

_____. "Mechanism, Purpose and Explanatory Exclusion," reprinted in *Supervenience and Mind*.

_____. "The Mind-Body Problem: Taking Stock After 40 Years," in *Philosophical Perspectives* 1997.

_____. *Mind in a Physical World*, MIT Press, Cambridge: 1998.

_____. "Multiple Realization and the Metaphysics of Reduction," reprinted in *Supervenience and Mind*.

_____. "The Myth of Nonreductive Physicalism," reprinted in *Supervenience and Mind*.

_____. "The Nonreductivist's Troubles with Mental Causation," reprinted in *Supervenience and Mind*.

_____."Psychophysical Laws," reprinted in *Supervenience and Mind.*

Lepore, Ernest and Barry Loewer. "Mind Matters," *Journal of Philosophy* 93 (1987): 630-642.

Leibniz, Gottfried Willhelm. *Discourse on Metaphysics,* trans. G.R. Montgomery (La Salle, Illinois: Open Court Publishing Co., 1927) reprinted in *18th Century Philosophy* ed. Lewis Beck (New York: The Free Press, 1966).

Levine, Joseph. "Materialism and Qualia: The Explanatory Gap," in *Pacific Philosophical Quarterly*, 64 (1983).

Loewer, Barry ed. *Fodor and His Critics*. (Blackwell: New York and Oxford, 1987).

Macdonald, Cynthia. "Tropes and Other Things," in *Contemporary Readings in the Foundations of Metaphysics*, edited by Stephen Laurence and Cynthia Macdonald (Blackwell: Massachusetts, 1998)

Macdonald, Cynthia and Macdonald, Graham. "Mental Causation and Non Reductive Monism," *Analysis*, vol. 50, pp.23-32.

Marras, Ausonio. "Nonreductive Materialism and Mental Causation," in *Canadian Journal of Philosophy*, vol. 24, number 3.

McLaughlin, Brian. "On Davidson's Response to the Charge of Epiphenomenalism," in *Mental Causation*.

Mellor, D.H. and Alex Oliver (eds.), *Properties*, (Oxford University Press: Oxford and New York, 1997)

Noordhof, Paul. "Do Tropes Resolve the Problem of Mental Causation?" in *American Philosophical Quarterly*, pp.221-228

Putnam, Hilary. "On Properties," reprinted in *Contemporary Readings in the Foundations of Metaphysics*, edited by Stephen Laurence and Cynthia Macdonald (Blackwell: Massachusetts, 1998)

_____. "The Nature of Mental States," reprinted in The Nature of Mind, edited by David Rosenthal (Oxford: University Press New York, 1992).

_____. "The Meaning of Meaning," in K. Gunderson, ed. *Language, Mind and Knowledge, Minnesota Studies in the Philosophy of Science*, 7 (University of Minnesota Press: Minneapolis, 1975).

Quine, Willard Van Orman. *Ontological Relativity and Other Essays*, (Columbia University Press: New York, 1968).

Robb, David. "The Properties of Mental Causation," in *The Philosophical Quarterly*, vol. 47, no. 187, April 1997.

Rosenthal, David ed. *The Nature of Mind* (Oxford University Press: New York and Oxford, 1991).

Russell, Bertrand. *The Problems of Philosophy* (Oxford University Press: London, 1959).

Searle, John R. *The Rediscovery of the Mind* (MIT Press: Cambridge, 1992).

Shoemaker, Sydney. "Causality and Properties," in *Properties*, D.H. Mellor and Alex Oliver (eds.) pp.228-255.

Sosa, Ernest. "Davidson's Thinking Causes," in *Mental Causation*.

Stich, Stephen P. *From Folk Psychology to Cognitive Science* (MIT Press: Cambridge, 1983).

_____. "Autonomous Psychology and the Belief Desire Thesis," reprinted in *The Nature of Mind* ed. by David Rosenthal.

Van Gulick, Robert. "Who's in Charge Here? And Who's Doing All the Work?, in *Mental Causation*.

Williams, Donald C. "On the Elements of Being: I," in *Properties*, D.H. Mellor and Alex Oliver (eds.) pp. 112-125

Yablo, Steven. "Mental Causation," in *Philosophical Review*, 101 (1992) pp.245-280.