This is to certify that the

dissertation entitled

Mental Causation and the Pragmatics of Explanation

presented by

Joseph Charles Totherow

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Philosophy

Richard J. Hall
Major professor

Date May 4, 1999

MENTAL CAUSATION AND THE PRAGMATICS OF EXPLANATION

By

Joseph Charles Totherow

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Philosophy

1999

ABSTRACT

MENTAL CAUSATION AND THE PRAGMATICS OF EXPLANATION

By

Joseph Charles Totherow

Most philosophers of mind presuppose or argue for some version of *physicalism*. However, against the backdrop of physicalism, there exists a fundamental tension between claims of the systematic irreducibility of the mental and claims that the mental, in virtue of its irreducible properties, is capable of causing behavior.

An implication of physicalism is that every behavioral event can, in principle, be given an exhaustive causal analysis or explanation through reference to bodily states and properties. Thus, to claim that irreducible mental states are causally efficacious is to claim that certain types of bodily behavior are subject to both an exhaustive physical explanation and an exhaustive mental explanation. This presents two problems, the solution to which is the main concern of this dissertation: First, how are we to make sense of the unique causal work performed by the mental if every behavioral event can be given an exhaustive causal analysis in the physical idiom? Second, if the mental does indeed perform unique causal work in the production of behavior, how are we to avoid the problem of systematic causal overdetermination?

This dissertation approaches these problems through a pragmatic theory of explanation, presupposing that the causal work of either mental or physical states can be understood in terms of the causal explanations that they provide. The salient feature of the pragmatic theory of explanation is that explanations are answers to why-questions. Thus, the unique causal work performed by irreducible mental states can be understood in terms of the causal explanations they provide. Specifically, the mental idiom can provide answers to particular why-questions that the physical idiom cannot. Further, causal overdetermination, according to the pragmatic theory of explanation, is understood as having multiple acceptable answers to a single why-question. Since the mental and physical idioms each provide answers to their own unique why-questions, the problem of causal overdetermination, on this analysis, does not arise.

Chapter I elaborates on the central problems of mental causation, explicating them according to the epistemological and metaphysical motivations of Jaegwon Kim's Explanatory Exclusion Principle (EEP). Chapter II develops a preliminary solution to these problems by responding to Kim's arguments. Chapter III develops certain details of the pragmatic theory of explanation, since it plays a vital role to the proposed solution. Chapter IV develops a more detailed solution to the problems of mental causation. Chapter V responds to the solutions offered by the eliminative materialist position and discusses the advantages of the present proposed solution.

TABLE OF CONTENTS

LIST OF FIGURES

INTRODUCTION

One of the truly classic problems in philosophy is the mind-body problem. But while philosophers have long grappled with the question of the nature of mind and the relation it bears to the body, the nature of the problem itself has changed dramatically over the past century.

A long-standing tradition in philosophy maintains not only a substantive difference between mind and body, but also a certain primacy of the mind over the body. As far back as Anaxagoras, we see this primacy of mind. For Anaxagoras, the origin of order, the first principle which set the physical universe in regular motion, is *Nous*, or Mind. Mind is therefore metaphysically prior to the physical world.

For Plato, the body is the cage of the immortal soul. The soul is the true nature of man, while the body serves only to obscure and blind the soul from its contemplation of the Forms. The soul, in so far as its proper place is in the heavens, contemplating the Forms, has a certain metaphysical priority over the body. The Forms are, after all, ultimately real.

Not quite so long ago, Descartes insisted that the mind, as a substance altogether different than physical substance, is better known than the body. Indeed, Descartes' epistemology begins not with an analysis of the knowledge of physical objects, but with the act of introspection on his

own faculty of thought. Thus, mind is epistemologically prior to body.

George Berkeley found the material world altogether expendable, retaining a metaphysics consisting *only* of minds and ideas. These few are but representatives of a long and pervasive view that the mind maintains an epistemological and metaphysical priority over the body. While there have been dissenters to this view throughout the history of philosophy, only within the past fifty to one hundred years has this conception of the mind fallen into general disrepute.

The reasons for this fall are numerous and complicated. For instance, by the mid-twentieth century, the claim that knowledge of the mind could be attained via introspection was criticized as inherently subjective and unreliable. The epistemological centrality of the mind would subsequently fall out of favor.

One might also claim that the emergence of pragmatism was a salient cause in the fall of the priority of mind. Rorty, for instance, exercising a fundamental tenant of pragmatism, rejects what he calls the "Platonic urge" to abstract oneself (i.e. one's mind) away from the physical world to the contemplation of Absolute Truth: "It is the impossible attempt to step outside our skins--the traditions, linguistic and other, within which we do our thinking and self-criticism--and compare ourselves with something absolute." (Rorty, xix)

Certainly, a thorough and detailed analysis of the fall of the primacy and substantive difference of the mind would be lengthy. But, in the broadest of strokes, it is reasonable to claim that the advent and growth of modern science is the primary cause. Since the revolution away from Scholastic science to the mechanistic world view of Galileo, both the scope and the explanatory and predictive power of modern science has grown exponentially. Along with this growth emerged a sort of positivism. Whatever could not be assimilated into this new scientific system must be dismissed as nonsense. E. A. Burtt gives an example of this positivistic attitude, drawn from Galileo's own works:

> In his discussion on the tides [Galileo] severely criticizes Kepler for explaining the moon's influence on the tides in terms that sound like the occult qualities of the scholastics, judging it better for people 'to pronounce that wise, ingenious, and modest sentence, 'I know it not' rather than suffer to escape from their mouths and pens all manner of extravagances.' (Burtt, 103).

In a world in which a (more or less) mechanistic, physical science dominates the intellectual landscape, the mind becomes an ethereal, alien entity. Analytic philosophers of mind of the twentieth century, by and large, are different from their predecessors in that a main goal of the discipline is to scientize the mind; to assimilate the mind into our scientific world-view. Somehow, our conception of the mind must find a place in a scientific world. But what impact would such an assimilation have on the idea that

the mind is epistemologically prior to the body and that, metaphysically, the mind is of a substance wholly different from the body?

If the above considerations are true, even in such broad terms, then the contemporary mind-body problem is fundamentally different from its traditional ancestor. While the traditional attempts to understand the nature of mind and the relation it bears to the body (such as Cartesian dualism) typically assume (or argue for) a difference in substance as well as a priority of mind, contemporary mind-body theorists have all but turned such assumptions on their heads. As such, contemporary theories of the mind have generally presupposed some version of *physicalism*.

While there have been different types of physicalism, usually varying with respect to strength, we can take Jaegwon Kim's notion of "minimal physicalism" to be representative of the general idea. According to Kim, there are two principles of minimal physicalism: "The anti-Cartesian principle", and "the mind-body dependence principle":

> [The anti-Cartesian principle] There can be no purely mental beings ... That is, nothing can have a mental property without having some physical property.
>
> [Mind-body dependence] What mental properties a given thing has depends on, and is determined by, what physical properties it has. That is to say, the

> psychological character of a thing is wholly determined by its physical character. (Kim II, 10-11)[1]

To assimilate the mind is to make it less alien to the physical world. Eliminating any primacy of the mental over the physical (the conditions of which are spelled out in the dependence principle) and precluding any version of substance dualism (a requirement met by the anti-Cartesian principle) goes a long way toward meeting this goal. Thus, contemporary philosophers of mind, if they are to take seriously the need to scientize the mind, or assimilate the mind into our scientific world-view, must at least construct their theories against the backdrop of minimal physicalism.

But what sorts of theories fit the constraints of minimal physicalism? What, properly speaking, does one study when one studies the mind? Is there a sense in which the mind is distinct from the brain, or, given the tenets of physicalism, must a study of the mind ultimately be a study of the brain?

Current discussions of the mind-body problem focus primarily on the relationship between physical states (primarily brain states) and mental states, e.g. qualia,

---

[1] Kim does invoke a third principle of minimal physicalism: The supervenience principle . This principle dictates that "any two things ... exactly alike in all physical properties cannot differ in respect of mental properties." However, supervenience is entailed by the dependence principle. While Kim distinguishes between supervenience and dependence (and lists them as separate principles), I shall regard them as a single principle, given that the latter entails the former.

beliefs, desires, etc.. Given that physicalism precludes substance dualism, it seems that the mind can be nothing over and above the brain. How, then, do we make sense of the relationship between mental and physical states?

Generally speaking, there are two main schools of thought on the matter. First, the ***reductionist*** argues that the mind can be systematically reduced to the brain. Thus, mental kinds can be identified with kinds of brain states. On such a theory, a necessary component of a scientific study of the mind is the discovery of the "bridge laws" by which mental states can be identified with brain states. Thus, a successful study of the mind must involve a systematic reduction of mental states to states of the brain, according to such bridge laws. For instance, in order to understand the mental state *pain*, we must understand its brain correlate, i.e. the firing of c-fibers.

Therefore, for the reductionist, there is no sense in which the mind is distinct from the brain. The language of the mind, i.e. talk of wants, beliefs, qualia, etc., is meaningful in that it refers to types of brain states. A thorough understanding and assimilation of the mind must therefore proceed according to a study of the brain.

On the other hand, ***functionalism*** argues that there *is* a sense in which the mind is distinct from the brain. For the functionalist, a successful assimilation of the mind into our scientific world view will not and cannot involve a type-reduction of mental states to brain states. Thus, the

functionalist must walk a metaphysical tightrope: There is a sense in which mental states are distinct from the brain, but they cannot be distinct in virtue of an instantiation in a second, non-physical substance; for this would clearly violate the principles of minimal physicalism. But, in what way can mental states be distinct without violating the tenets of physicalism?

The functionalist claims that we can understand and define mental states functionally. In other words, we can understand a particular mental state in terms of its causal role in a cognitive system, where that role consists of relations to input stimuli, to other mental states, and to output behavior. For instance, the mental state *pain* can be understood in terms of its causal relations to input stimuli (stubbing one's toe), to other mental states (the desire to cry out, the desire to protect the wounded area, etc.) and to output behavior (crying out, rubbing one's toe, swearing, etc.). On this view, a study of the mind is a study of the abstract causal network of mental states, as well as the varieties of input stimuli and output behavior. Mental states are therefore, on this view, functional states.

Thus, at this point, two questions arise: First, is a functional understanding of mental states necessarily incompatible with the reductionist's identification of mental states with brain states? Why is it that we cannot conceive of *pain* both functionally and in terms of a certain physical kind? Second, if they are indeed mutually exclusive, which

analysis of the mental is superior? Are there any arguments to recommend the reductionist's analysis over the functionalist's, or vice versa?

First, the functionalist argues that a functional understanding of mental states is incompatible with a reductionist conception of the mind. To conceive of mental states as functional states is to claim that they are not systematically identifiable with brain states--that the mind is somehow distinct from the brain. The most influential argument for this claim is the *multiple realizability argument*.

Suppose Jones is a mechanic and Smith is a mechanically-illiterate friend. Smith asks Jones to explain what a carburetor is. Jones has (at least) two choices for her explanation: Either she explains what a carburetor is according to a functional analysis or according to a reductionist analysis. On a functional analysis, Jones would explain the causal role of the carburetor relative to the rest of the car: The carburetor responds to the driver's pressing on the accelerator (external stimulus) by increasing the amount of oxygen and fuel sent to the engine cylinders (other internal states of the system), which turns the gears of the transmission more quickly, thereby increasing the rate of tire rotation and accelerating the automobile (behavioral output). On a reductionist analysis, on the other hand, Jones would explain the concept of the carburetor in terms of

its physical constitution, e.g. the material of which it is made, namely a certain type of metal alloy.

Again, the functionalist claims that the functional conception of the carburetor is incompatible with the reductionist conception. This is because the carburetor, functionally defined, can be realized in a number of different physical types. While Jones' carburetor happens to be made of a particular metal alloy, it could very well have been made of a type of hardened fiberglass and still performed its causal role, in relation to the rest of the automotive system. Thus, since the carburetor, functionally conceived, is multiply realizable in an indefinite number of physical types, the reductionist analysis of the carburetor as identical to a particular physical type is incompatible with the functional analysis.

But while the functionalist and reductionist analyses of the carburetor are incompatible, it is not yet clear whether one is superior to the other. In order to distinguish the superior of the two analyses, we need to ascertain which understanding more closely resembles our common conception of the carburetor. Would a close analysis of our ordinary, shared definition of the carburetor more closely resemble the reductionist conception or that of the functionalist?

Again, the functionalist turns to the multiple realizability argument. Let us return to Jones and her automobile. Jones' carburetor has broken down and must be replaced. However, while Jones has the moldings for shaping

a new carburetor, she does not have a supply of the metal alloy she needs to fashion a new metal carburetor.

Instead, she uses her moldings to fashion a device out of a super-hardened fiberglass. Once cooled, she installs this device in her car where the carburetor used to be. Once installed, Jones turns the key and the car starts and continues to run. Functionally speaking, this new fiberglass piece of machinery bears the same causal relations to the rest of the car as did her old carburetor. In other words, the car with the fiberglass mechanism is *functionally isomorphic* to the car with the metal carburetor. The question is, is this new fiberglass mechanism, strictly speaking, a carburetor? For instance, would the community of mechanics recognize Jones' machine as a carburetor?

The functionalist predicts (rightly, I think) that the new mechanism, regardless of its material, will be considered by anyone who understands automotive systems to be a carburetor. The functionalist argues that the functional definition of the carburetor is not only incompatible with the reductionist's, it is superior, because it more closely resembles our common idea or conception of what a carburetor is. The fact that a mechanic would define Jones' new fiberglass mechanism as a carburetor illustrates that our common definition of a carburetor is a functional definition.

But what of mental states? First, are mental states multiply realizable? Second, would a functional definition of mental states be superior to the reductionist definition?

The multiple realizability of mental states can be established by a thought experiment similar to our carburetor story.

Let us take a common example: An alien, whose chemical composition is based in silicon (whereas humans' is based in carbon), lands its spacecraft just outside of Lansing. First, while the alien is composed of a radically different physical type than humans, it is perfectly plausible that it maintains internal states that, functionally defined, mirror our own. As Paul Churchland says,

> The chemistry and even the physical structure of the alien's brain would have to be systematically different from ours. But even so, that alien brain could well sustain a functional economy of internal states whose mutual relations parallel perfectly the mutual relations that define our own mental states. (Paul Churchland, 36-37)

Thus, it is quite plausible that such an alien could admit of internal states that mirror (or are *functionally isomorphic* to) our own.

Therefore, an alien, quite unlike us in physical makeup, could plausibly maintain internal states that admit of functional definition. So, for each such state, there is a functional definition and a reductionist definition. Again, the question arises, which definition is superior?

Let us say that the alien begins to advance towards a group of people who have gathered to watch the ship's landing. As it advances, it extends an appendage towards the

group. Panicked, a member of the group clubs the appendage with a nearby branch. Once struck, the alien shrieks, retracts the appendage, shields it from further damage and moves quickly back into the spaceship. Would we want to claim that the alien felt pain? The functionalist reasonably predicts that, were such a situation to occur, the internal state *pain* would be attributed to the alien, even though it is made of a radically different type of substance. The alien felt pain and wished to protect the wounded area and avoid any further infliction of injury. The reductionist, on the other hand, cannot say that the alien felt pain because it has no c-fibers.

As was the case with the carburetor example, the functional definition of *pain* is superior to the reductionist definition because the former more closely mirrors the common use of the concept. The functionalist predicts that, if such a being were to land on earth, we would, as a community of speakers who maintain a mental idiom, attribute mental states to it. Thus, not only are the functionalist and reductionist analyses of the mental incompatible, the functionalist analysis of mental states is superior in that it more accurately captures our common use of mental concepts.

Hence, the functionalist is able to walk the metaphysical tightrope: Functionalism is able to maintain the irreducibility of the mind by denying type-identifications of mental states with physical states. Thus, the mind is distinct from the body, but not in virtue of

being a second, immaterial substance. On a functional conception of the mind, mental states must be realized in a physical system. But it does not follow that the mind simply *is* the physical system that realizes it. Thus, humans are physical beings, but with minds that are not systematically reducible to their brains.

The primary virtue of functionalism is its ability to make sense of mental states as distinct from brain states without violating the principles of minimal physicalism. Thus, in terms of the question of the assimilation of the mind into our scientific world view, it appears that functionalism is superior to Cartesian dualism. But while it avoids the problem of invoking a second immaterial substance, functionalism, on its own, does not escape all the difficulties that afflict Cartesian dualism.

A primary example is the problem of mental causation. The mind, it is usually believed, is capable of having a causal impact on certain types of human (and possibly animal) behavior. At bedtime, a child may ask a parent to leave the bedroom light on *because* he is afraid of the dark. If one cries, it is *because* he is sad. If I go to the store, I do so *because* I believe I am out of milk and I believe that, come tomorrow, I will desire that I have milk with my breakfast. Fear, sadness, beliefs and desires are mental states and it is not uncommon to believe that such mental states affect behavior.

Further, the causal efficacy of the mind is a necessary presupposition in many attempts to solve the problem of free will. C.A. Campbell, for instance, in his notion of phenomenological analysis, presupposes that deliberation, as a subjective state of the mind, has a causal impact on subsequent behavior[2].

Finally, it would seem obvious that claims of moral responsibility for one's actions, as they are judged in courts of law, must necessarily involve reference to a defendant's state of mind. This presupposes that one's state of mind can impact subsequent behavior. Thus, in the development of a theory of the mind-body relation which makes the mind assimilable into our scientific world-view, a theory of the irreducibility of mental phenomena can only be part of the story. An account of the causal efficacy (or the lack thereof) of the mental must also be provided. It is with the question of mental causation that functionalism finds itself in a bind similar to that of its Cartesian ancestor.

Dualism's problem with mental causation seems intractable and historically is one of the difficulties which ultimately led to its current unpopularity. With dualism, one finds two disparate types of substances which share no common attributes. On such a theory of the mind-body relation, how would a theory of mental causation work?

---

[2] See Campbell, On Selfhood and Godhood, (London: George Allen & Unwin) 1957, pp. 158-179.

For dualism, physical states, such as stubbing one's toe (P1), cause mental states, such as the feeling of pain (M1). Further, mental states, such as the desire to ease the pain and the belief that rubbing the toe will ease the pain (M2), could cause physical actions, such as rubbing the wounded toe (P2).

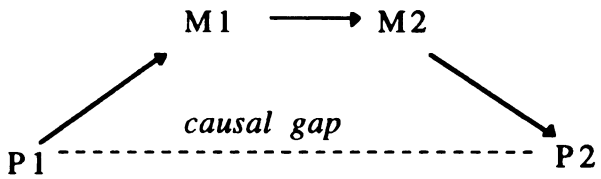

Figure 1

But certainly, on the dualist account, this would run contrary to the causal closure of the physical universe and to the law of the conservation of energy. If this characterization of the toe-stubbing incident is correct, then there is a "causal gap" at the physical level between the stubbing and the rubbing of the toe. On a physical analysis, new energy or work is introduced into the physical universe, since the rubbing of the toe has no direct physical cause. This certainly runs counter to the laws of the physical sciences.

The causal gap problem illustrates the difficulty of assimilating Cartesian dualism into our scientific world-view. Indeed, it is reasonable to claim that this problem is a primary motivation for the "anti-Cartesian principle" of minimal physicalism. However, banishing the second substance

does not automatically resolve the problems of mental causation. As will be shown, functionalism, even without invoking a second substance, is afflicted with difficulties strikingly similar to the causal problems of dualism. It too finds difficulty assimilating into our scientific world-view.

Again, the primary virtue of functionalism is its ability to construct an understanding of the irreducibility of the mental without invoking a second substance. Thus, while the mind cannot be systematically reduced to a physical system (in our case, the brain), it cannot exist unless instantiated in a physical system. The functionalist is able to walk this tightrope by invoking the *type/token distinction*.

A systematic reduction of mental states to physical states would involve a type-identification of mental and physical states. Bridge laws produce this sort of type-identification. The functionalist analysis precludes such a type-identification, but insists that each particular (token) mental state must be instantiated in a physical medium. Mental states are token-identical to physical states, but not type-identical.

In virtue of the type/token distinction, functionalism is able to avoid the "causal gap" problem of substance dualism. Since mental states are token-identical with physical states, maintaining a causal relationship between mental states and physical states does not imply a causal gap

at the physical level, at least at first blush. According to the functionalist, a mental state is causally efficacious only if it is realized in a physical system. Roughly, mental states inherit their causal powers from the physical states in which they are realized. Maintaining the causal efficacy of the mental, on this view, does not preclude the causal continuity between physical events. However, even if the causal gap problem is avoided, functionalism is plagued with its own unique problem with mental causation.

The new problem of mental causation is given voice by Jaegwon Kim. Kim argues that the functionalist's attempt to account for mental causation turns on the phrase "determined by but not identical to." (Kim I, 355) To put this phrase in context, the causal powers of mental states are "determined by but not identical to" the causal powers of physical states. So as to avoid the Cartesian problem of mental causation, mental states must inherit their causal efficacy from physical states. But in order to retain the irreducibility, as well as the causal efficacy of the mental, the causal potency of mental states cannot be identified with that of the physical.

One could easily imagine a functionalist version of the "determined by but not identical to" (hereafter, DBNI) thesis. Mental states must be realized in a physical medium (via token-identity). Thus, the causal efficacy of mental states is "determined by" the causal powers of the corresponding physical states. However, given that those

mental states (including their causal properties) can be instantiated in any number of physical types, neither the mental states nor their causal properties can be "identified with" any particular type of physical state. But, as Kim shows, such an account of mental causation leads to serious difficulties[3].

Suppose mental state M is realized in physical state P. Further, M is the cause of a second physical state, P*. More precisely, M is a sufficient cause of P*. According to the DBNI thesis, M inherits its causal powers from its physical realization, namely P. In other words, M is able to cause P* because M is realized in P. But, also according to the DBNI thesis, the causal properties of M are not identical to those of P. Thus, M causes P* in virtue of novel causal properties belonging to M.

Now, it would be highly problematic to claim that, while M is a sufficient cause of P*, P is not also a sufficient cause of P*. Indeed, such a claim would run counter to the DBNI thesis. Again, the thesis states that the causal efficacy of a mental state is determined by its physical realization. But if M were a sufficient cause of P* and P were *not* a sufficient cause of P*, then the causal efficacy of M would *not* be wholly determined by P, since M would have

---

[3] For a full explication of the following argument, see Kim's "The Non-Reductivist's Troubles with Mental Causation", sec. VI, in Supervenience and Mind, (1993).

a causal potency that P lacks, namely, being sufficient for P*. This would render the DBNI thesis false.

Thus, it follows from the DBNI thesis that, where M and P are token-identical (but not type-identical) and P* is a state subsequent to M and P, M is a sufficient cause of P* if and only if P is a sufficient cause of P* (see fig. 2). As Kim states it, "since P is a realization base for M, it is sufficient for M, and it follows that P is sufficient, as a matter of law, for P*" (Kim I, 354).

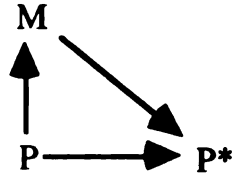

Figure 2

Here we come to the heart of the problem as Kim sees it. The irreducibility of M to P, in conjunction with maintaining the causal efficacy of M, leads to *causal overdetermination*:

> We would be allowing two distinct sufficient causes, simultaneous with each other, of a single event. This makes the situation look like one of causal overdetermination, which is absurd. ... Given that P is a sufficient physical cause of P*, how could M *also* be a cause, a sufficient one at that, of P*? What causal work is left over for M, or any other mental property, to do? (ibid. 354)

Hence, the similarity between Cartesian dualism and functionalism becomes readily apparent. Both theories appear

to resist the assimilation of causally efficacious mental states into our scientific world view. On the one hand, dualism violates the principles of the conservation of energy and the causal closure of the physical universe. On the other hand, the functionalist analysis of mental causation apparently involves causal overdetermination.

But what's wrong with causal overdetermination? Certainly, there is the occasional causally overdetermined event, such as a firing squad in which a man is shot in the heart with two bullets at the same instant, each bullet being sufficient for his death. Why should this pose such a problem for mental causation?

The problem is, claims of mental causation (either from dualism or functionalism) do not involve the occasional, anomalous case of overdetermination. Rather, mental causation involves *systematic overdetermination*, where every intentional or purposive behavioral event is causally overdetermined. Systematic overdetermination occurs when our theories of the world (in the present case, our theories of the causal efficacy of the mind and that of the body) entail overdetermination.

Kim claims that at best, systematic causal overdetermination is "very odd" and at worst, "absurd." He acknowledges that causal overdetermination does occur. But it is difficult to give a causal explanation in such cases. Indeed, if two distinct, causally unrelated events M and P are each sufficient for the production of event P*, then it

is unclear how we should formulate a causal explanation of P*. Both M and P render each other unnecessary in the explanation of P*. So in general, genuine cases of overdetermination make explanation difficult.

The problem for mental causation goes deeper than merely presenting difficulties for explanation. As Kim says, "it is at best extremely odd to think that each and every bit of action we perform is overdetermined in virtue of having two distinct sufficient causes." (Kim I, 247) Cases of overdetermination, generally, are oddities-- explanatory anomalies, if you will. It would be "extremely odd" to have our view of the world set up to claim that every bit of intentional behavior (which encompasses quite a bit of bodily movement) is causally overdetermined. Such a claim would be akin to having a thorough understanding of the buildup and discharge of static electricity in the atmosphere as an explanation for lightning, as well as the theory that each lightning bolt is hurled to earth by Zeus because he is angry. (Whether or not common-sense psychology is on a par with Greek mythology is a matter of some debate in contemporary mind theory.) It appears that this is what Kim means by calling systematic overdetermination "absurd."

It seems that accounting for the irreducibility of the mental without invoking a second substance does not solve all the difficulties packed into the mind-body problem. Indeed, against the backdrop of physicalism, the problem of mental

causation is as much a problem for functionalism as it is for Cartesian dualism.

It is the goal of this dissertation to present an account of mental causation which, within the conditions of physicalism, is reconcilable with the irreducibility of the mental. Such an account requires, first, a closer analysis of Kim's causal-overdetermination argument. This argument, as it applies to mental causation, is found in the more general argument for, what Kim calls, the "explanatory exclusion principle" (hereafter, EEP). Thus, the resolution of the functionalist problem of mental causation will begin with a response to Kim's general arguments for explanatory exclusion.

Chapter I: Mental Causation and Explanatory Exclusion

1. The Explanatory Exclusion Principle

A rather stock example, first given by Norman Malcom (1968) and subsequently employed by Kim himself, will help facilitate the discussion of the EEP: As Smith walks to her car, a gust of wind blows her hat to the roof of her garage. She ponders the situation briefly, then finds a step-ladder in the garage, climbs it and retrieves her hat. While such a task may seem rather mundane, the cognitive skills involved in Smith's solving the problem are actually rather impressive. While this behavioral series is complex, for the sake of argument, the series will be considered as a single event, E. E, like other non-quantum physical events, is subject to a causal explanation. But what is that explanation? Where would one look for such an explanation?

E appears subject to at least two types of explanation. First, one might give an explanation X which is formulated in the intentional idiom: Smith wanted to retrieve her hat; she believed that she could not reach the hat without aid; she desired to have such an aid that would assist her in the retrieval of her hat; she believed there was a ladder in the garage, etc. Second, an explanation Y, couched in the language of neurophysiology, can be offered: Certain external stimuli caused a particular series of neurons to fire in Smith's brain, which, in turn, caused a certain

electro-chemical reaction, which sent various impulses down her spinal cord to her limbs, which caused her muscles to contract in a certain order, etc.

Kim argues that such a situation, where there are multiple explanations of a particular event, *may* be untenable. The conditions under which such a situation is untenable are expressed in what Kim calls "the explanatory exclusion principle": "No event can be given more than one *complete* and *independent* explanation" (Kim I, 239). As applied to the example, X and Y cannot both be complete and independent explanations of E. In order to begin to understand the meaning and importance of the explanatory exclusion principle, it is necessary to understand Kim's conception of the nature of explanation in general.

Kim makes two explicit presuppositions about explanation. First, he assumes a strong, symmetrical relation between causation and explanation, such that X is a cause of Y if and only if X is an explanation of Y[4]. A second, related presupposition is what Kim calls "explanatory realism": "That a causal explanation of E in terms of C is a 'correct explanation' only if C is in reality a cause of E, can be called an 'objectivist' or 'realist' conception of explanation" (Kim, 256). Thus, an explanation of an event E

---

[4] While this construal of causation, explanation and their relationship may seem overly strong, I will not dispute this premise of Kim's argument, as I think that he is, in general, correct.

is correct (true, accurate, or what have you) in virtue of picking out the actual cause which brought about event E.

At the heart of the exclusion of multiple explanations are the criteria of completeness and independence. Therefore, in order to understand the EEP, we must understand the meaning of these criteria. Curiously, Kim never gives an explicit definition of either concept. Instead, an almost purely negative definition arises from examples or cases Kim gives in which explanations are either dependent, incomplete or both. While this poses some difficulty in divining Kim's exact meaning of these criteria, we should be able to generate a loose but positive conception of completeness and independence which will be adequate for our purposes.

First, in situations where X and Y only *seem* to be distinct, but are actually type identical, neither X nor Y is an independent explanation of E. Such would be the case if, for example, there existed bridge-laws that allowed scientists to type-reduce biological concepts to those of chemistry. Explanations which invoke biological concepts would be dependent on chemical explanations. Further, if the events described by X supervene on those specified by Y, then X is not independent of Y as an explanation. Also, where X and Y are both links in the same causal chain, neither X nor Y is an independent explanation.

Regarding completeness, Kim argues that if X is a "proper part" of Y (i.e. if X picks out a proper part of the causal process specified in Y), then X is neither complete

nor independent. Finally, Kim does give at least a partial positive definition of completeness: "In one sense of complete, ... a complete explanation specifies a sufficient set of causal conditions for the explanandum" (Kim, 251).

These clues generate a fairly definite conception of completeness and independence. Independence appears to be a relational term: To say that explanation X is independent is to say that X is independent of other possible explanations, i.e. X neither collapses into (via an identity relationship) nor supervenes over another explanation Y, nor is it a proper part of another explanation. Thus, for X to be independent as an explanation, X must pick out a causal process which relies upon no other causal process for the production of E.

And, for X to be a complete explanation, X must exhaustively specify each component of a particular causal series, having E as its result, which makes that causal series a sufficient condition for E. What Kim seems to have in mind by these criteria, then, is that X is a complete and independent explanation of E if and only if X exhaustively picks out a causal process which produces E (i.e. X specifies events X1 ... Xn which are jointly sufficient for the production of E), without bearing a dependence relation (supervenience, identity, and the like) to any other causal process.

Now that the concepts of completeness and independence are more or less clear, why can there not be more than one complete and independent explanation of a particular event?

Recall Smith's hat-retrieval behavior. At least two explanations of her behavior are forthcoming-- X, couched in the intentional idiom and Y, expressed in neurophysiological terms. If both X and Y are complete and independent explanations of E, then each refers to a causal series which is sufficient for the production of E. Further, neither explanation can somehow be reduced to the other, via supervenience or identity. Roughly, there are two distinct causal processes, both of which are responsible for the occurrence of E.

According to Kim, this situation is untenable. He first objects to the possibility of such a situation on metaphysical grounds. The immediate problem, again, is causal overdetermination, which, as Kim says, is a "metaphysically unstable" situation. Recall that while there are instances of causal overdetermination, they are explanatory anomalies. They are situations which make explanation difficult:

> A man is shot dead by two assassins whose bullets hit him at the same time; or a building catches fire because of a short circuit in the faulty wiring and a bolt of lightning that hits the building at the same instant. It isn't obvious in cases like these just how we should formulate an explanation of why or how the overdetermined event came about. (Kim II, 252)

The problem of overdetermination (the "metaphysical instability") emerges when we conceive of the world as systematically overdetermined. Mental causation illustrates this problem nicely. A conception of mental causation which

maintains the irreducibility of the mental violates the EEP on metaphysical grounds, since it conceives of the world as systematically overdetermined. It makes every piece of intentional bodily behavior an explanatory oddity or anomaly. It appears that the "absurdity" of systematic overdetermination is the basis for this metaphysical instability. It is bad enough that the occasional bona fide case of causal overdetermination baffles our attempts at explanation, since each explanation renders the other (or others) superfluous, but to set up our view of the world such that all cases of purposive action are causally overdetermined is absurd.

Kim also claims that having multiple complete and independent explanations of an event is an epistemically unstable or counterproductive situation: "When we look for an explanation of an event, we are typically in a state of puzzlement, a kind of epistemic predicament. A successful explanation will get us out of this state" (ibid. 254).

Thus, a fundamental epistemic principle of explanation is that a satisfactory explanation must alleviate puzzlement. It is clear that if either X or Y were exclusive explanations of E, then our puzzlement would vanish. However, where we have both X and Y as complete and independent explanations, our puzzlement is not alleviated: Our old epistemic predicament of needing an explanation for E is compounded by the new epistemic predicament of needing to know which explanation is correct. As Kim says, "too many explanations

will put us right back into a similar epistemic predicament" (ibid. 254). Having multiple competing explanations is counterproductive to the epistemic goals of explanation, i.e. removing this type of epistemic puzzlement.

So there are (at least) two motivations for the explanatory exclusion principle: A metaphysical quandary, where we are faced with the absurdity of systematic causal overdetermination, and an epistemic quandary, where having multiple explanations is counterproductive to the epistemic goals of explanation. Based on these difficulties, Kim claims that we cannot maintain multiple complete and independent explanations of a particular event. All but one explanation-candidate must be either incomplete, dependent, or both.

How does the EEP apply to Smith's hat-retrieval behavior? Since the intentional explanation X and the neurophysiological explanation Y cannot both be complete and independent, we must exclude either one or the other (or both) as a complete and independent explanation of E. Certainly, it would be exceedingly problematic to maintain that the physical idiom cannot provide a complete explanation. Further, it would be even more difficult to maintain a dependence of physical explanations upon mental explanations.

In the first place, Kim cites Churchland's arguments[5]

---

[5] From "Eliminative Materialism and the Propositional Attitudes", *Journal of Philosophy* 78 (1981), p. 73.

regarding the explanatory failure of the intentional idiom: "As examples of central and important mental phenomena that remain largely or wholly mysterious within the framework of [folk psychology], consider ... the faculty of creative imagination ... the common ability to catch a fly ball ... the miracle of memory..." (Kim I, 262).

Churchland cites further examples of the intentional language's (or, as he says, "folk psychology's") inability to explain various types of psychological phenomena: Within folk psychology, one cannot explain the need for rest, the different levels of intelligence displayed in the human population, nor are we able to explain or prescribe cures for mental illness (Churchland, 45-46). Such examples prove difficult for those who claim that physical explanations depend on mental explanations, since in these cases, there is no mental explanation on which the physical explanation might depend[6].

Explanatory failure of the mental idiom is not the only problem with maintaining a dependence of the physical on the mental. Indeed, maintaining such a dependence would amount to a return to Descartes' "causal gap" problem. Recall the

---

[6] These examples seem to preclude a dependence of physical explanation on mental explanation. But while Kim uses these examples to argue for a dependence of the mental on the physical, Churchland uses them to advance his theory of *eliminative materialism*. This theory predicts that intentionality, as a means of describing human cognition, will ultimately be replaced by a fully-developed theory of neurophysiology. While the immediate concern is to address Kim's arguments for the exclusion of mental explanation, Churchland's arguments shall be addressed in chapter V.

"toe-stubbing" example. Stubbing one's toe causes (at least) pain and the desire to alleviate that pain, which, in turn, causes crying, swearing and the rubbing of the wounded toe. Both the toe-stubbing and the subsequent behavior are physical events. According to the neurophysiologist, the stubbing causes certain impulses to be sent to the brain, which in turn, causes the subsequent physical behavior.

To maintain the dependence of this explanation on the mental explanation entails that the toe-stubbing and the subsequent brain events are causally connected only in virtue of the brain events' dependence on the mental events caused by the toe-stubbing (i.e. pain and the desire to end the pain):

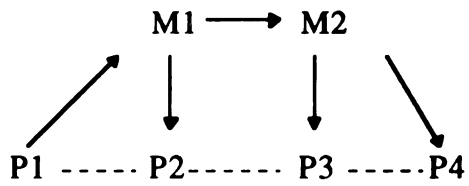

Figure 1

In such a situation, where the physical explanation is dependent upon the complete and independent mental explanation, the "real" causal series is P_1 -> M_1 -> M_2 -> P_4. If the physical series P_1, P_2, P_3, P_4 (where P_1 is the toe-stubbing and P_4 is the swearing and rubbing) is to be a causal series, it must be a causal series in virtue of its dependence on the former. Such a dependence entails causal gaps between physical events. This brings us back to the

"causal gap" problem of Cartesian dualism. Thus, even without the problem of explanatory failure, maintaining the explanatory dependence of the physical on the mental runs counter to our scientific world-view in exactly the same manner in which substance dualism does. Unless one wishes to return to the problems of Cartesian dualism, maintaining the dependence of physical explanations on mental explanations is out of the question.

The most satisfactory application of the EEP, then, is to maintain a dependence of mental explanations on physical explanations. There are at least two accounts of such a dependence relation: Either mental states are dependent in virtue of being type-identical with physical states or in virtue of an epiphenomenal relation with physical states. The first option is ruled out by the irreducibility of the mental. Therefore, if Kim's arguments about explanatory exclusion are correct, the most adequate anti-reductionist conception of mental causation is an epiphenomenal relation, in which mental states are only causally efficacious in virtue of their physical instantiations. As far as mental causation is concerned, irreducible mental properties there may be, but they are epiphenomena and can have no direct impact on physical behavior.

6. Mental Causation and the Levels of Analysis

But, one might argue, doesn't this stand to reason? Causality--the real "pushes and pulls"--takes place at the physical level. Mental states, by functional definition, are not physical (at least at the level of types). Thus, mental states can only be causally efficacious if they are instantiated in a physical body.

The functionalist, in keeping with the conditions of physicalism, concedes that the mind is causally efficacious only if physically realized, but it does not logically follow that the mind is causally efficacious *in virtue of* that physical realization. Recall that, for the functionalist, mental states are type-irreducible to physical states since the former can be realized in a number of physical types. Thus, no functional property is, properly speaking, a property of the physical medium, since that functional property can be realized in other physical types.

However, if the EEP is correct and the mental is causally efficacious *only* in virtue of its physical realization, then causal properties are not proper to functional states. Causal properties are analyzed at the physical level. The intentional language, therefore, can figure only indirectly in questions of causality. Therefore, *if the EEP is correct, then if we maintain the irreducibility of the mental, via functionalism, the mental is stripped of its causal efficacy.* Not only do mental states supervene

over physical states, but causal properties can only be located at the subvenient (i.e. physical) level. Properly speaking, there are no causal mental properties.

So while philosophers of mind have made great strides in maintaining the irreducibility of the mental without violating the principles of physicalism, it appears that there is no room for both an irreducible and a causally efficacious conception of the mind within a physicalist framework. This brings us to the central issue: Is it possible to construct a positive conception of mental causation which can be reconciled with the irreducibility of the mind, without violating the principles of physicalism? If such a project were successful, the more general project of assimilating the mind into our scientific world-view will have taken a significant step forward. It is this project that is the main concern of this dissertation.

It is important to note that the tension between mental irreducibility and mental causality (a tension which manifests itself relative to the conditions of physicalism) is representative of a larger problem. As Fodor points out in "Making Mind Matter More"[7], to question the causal efficacy of the mental is to question the causal efficacy of the properties of the special sciences in general:

---

[7] From A Theory of Content and Other Essays (1990), pp. 137-160.

> If there's a case for epiphenomenalism in respect of psychological properties, then there is the same case for epiphenomenalism in respect of all the nonphysical properties mentioned in theories in the special sciences. (140)

Thus, there is more at stake here than simply a conception of mental causation. If the constraints of physicalism apply to mental states and properties, then it must apply to the properties of all the special sciences. If the EEP is truly a reasonable constraint on explanation in general, then it must be applied across the board to all the special sciences. It is hoped that the present proposal for the reconciliation of the irreducibility and the causal efficacy of the mental will provide a model for understanding the causal efficacy of the states or entities of other special sciences.

Chapter II: Explanatory Exclusion and the Pragmatics of Explanation

1. Explanatory Exclusion and the Non-basic Sciences

The scope of the explanatory exclusion principle goes well beyond considerations of the causal properties of mental states. There are a number of non-physical sciences (i.e. those sciences which maintain a vocabulary and ontology beyond that invoked by physics) that posit the existence of states and objects which are instantiated in, but resist systematic reduction to, physically described states and objects. At first blush, the EEP's broad scope might seem to present serious problems for a wide variety of non-physical or special sciences. Ironically, however, it is the scope of its application that makes the EEP highly dubious.

If considerations of the application of the explanatory exclusion principle are limited to mental causation, then Kim's arguments might appear to make sense. It does not require a tremendous stretch of one's intuitions or imagination to exclude mental states as causally efficacious, outside of their physical realizations. After all, the mind is an ethereal, mysterious thing, even without the invocation of a second substance.

But if the EEP is applied to other non-physical sciences it loses a good deal of its original luster. Certainly, there are sciences which resist systematic reduction--usually

by some multiple-realizability argument--but whose states and properties are clearly causally efficacious. Michael Scriven, in "Causation as Explanation", gives examples of the indispensability of causation in the economic sciences: "The debate over the cause of economic events such as the present inflation/depression is one in which the eminent economists involved will not accept translations of the dispute into non-causal (e.g. correlation) terms." (Scriven, 4)

Jerry Fodor, in "Making Mind Matter More," is quite explicit about the implications of stripping the mental of its unique causal properties: "If there's a case for epiphenomenalism in respect of psychological properties, then there is the same case for epiphenomenalism in respect of *all* the nonphysical properties mentioned in theories of the special sciences." (Fodor, 140) Fodor gives several examples to illustrate this point, including the science of geology, which maintains the property of "mountainhood":

> Consider, for example, the property of being a mountain; and suppose (what is surely plausible) that being a mountain isn't a physical property ... Untutored intuition might suggest that many of the effects of mountains are attributable to their *being* mountains ... It is because Mount Everest is a mountain that Mount Everest has glaciers on its top; ... that it casts such a long shadow ... and so on. (139)

But geology appears to resist systematic reduction to the physical sciences; for if any non-physical properties are to be considered multiply realizable, "mountainhood" must be counted in their number. Without such a systematic

reduction, however, the attribution of causal efficacy to the states and properties of geology must run counter to the explanatory exclusion principle. Thus, for reasons similar to those given for the exclusion of mental explanation, the science of geology must be understood as causally inert.

A second example given by Fodor is the airfoil, a concept contained in the science of aerodynamics. Again, the claim is that the airfoil is multiply realizable. Thus, the properties of the airfoil cannot be type-reduced to properties described in the physical idiom. However, it would seem absurd to claim that the properties of being an airfoil are causally inert:

> Typically, airfoils generate lift in a direction, and in amounts, that is determined by their geometry, their rigidity, and many, many details of their relations to the (liquid or gaseous) medium through which they move. The basic ideas is that lift is propagated at right angles to the surface of the airfoil along which the medium flows fastest, and is proportional to the relative velocity of the flow. (139)

Ceteris paribus laws regarding the propagation of lift by an airfoil go a long way towards explaining why planes are able to fly and why sailboats are able to move through the water. Such explanations, it would seem, are causal. But the application of the EEP precludes such causal explanations. For reasons similar to those given in chapter I for the exclusion of mental explanations, we must exclude aerodynamic explanations: They are subject to explanatory failure, (their laws are, after all, ceteris paribus laws) and their

scope is more limited than that of the physical sciences. However, to apply the EEP in this case and preclude the causal efficacy of airfoils is, according to Fodor, "quite mad": "Airplanes fall down when you take their wings off; and sailboats come to a stop when you take down their sails." (140)

In the search for causal analyses of the movement of sailboats or the flight of airplanes, causal claims and explanations which refer to the properties of airfoils would be appropriate. As Fodor argues, "if that isn't the right explanation, what keeps the plane up? If that is the right explanation, how could it be that being an airfoil is causally inert?" (140) Therefore, as with the case of geology, an application of the explanatory exclusion principle to the properties of airfoils yields absurd results.

Further examples could be given, but the point is clear: If we accept the explanatory exclusion principle as a "plausible constraint on explanation in general" (Kim I, 239), then none of the non-basic sciences are capable of making causal claims or generating causal explanations. As far as Fodor is concerned, such implications make the arguments for the epiphenomenalism of the mental at best, suspect, and at worst, absurd:

> There are lots ... of examples where, on the one hand, considerations like multiple realizability make it implausible that a certain property is expressible in

> the physical vocabulary; and, on the other hand, claims for the causal inertness of the property appear to be wildly implausible. (139)
>
> Though it's true that claims for the epiphenomenality of mountainhood and airfoilhood and, in general, of any nonphysical-property-you-like-hood, will follow from the same sorts of arguments that imply claims for the epiphenomenality of beliefhood and desirehood, it's also true that such claims are prima facie absurd. (141)

The far-reaching implications of the EEP appear to be its undoing. Examples such as those given by Scriven and Fodor show that the explanatory exclusion principle is not, as Kim maintains, "a plausible constraint on explanations in general." Rather, it is simply too strong a condition.

However, the problems and considerations which motivate the EEP remain steadfast: How are Kim's metaphysical and epistemological quandaries regarding multiple explanations to be addressed? More broadly, if the EEP is incorrect, how are cases of multiple complete and independent explanations to be understood? How are we to make sense of the relationship between such explanations?

2. The Friendly Physicist and the Multiplicity of Explanations

In order to answer these questions and, in general, understand the problems that make the EEP untenable, it is necessary to begin with a closer analysis of an example (similar to those given by Fodor and Scriven) which appears to run afoul of the EEP. By now, most philosophers have heard Wesley Salmon's story of the "friendly physicist." A

physicist on an airplane, waiting to take off, makes a friendly wager with a young boy seated nearby: He contends that the boy's helium-filled balloon, on takeoff, will move toward the front of the plane. The boy (along with the other passengers) believes that, just as a passenger feels the inertia pulling him toward the back of the plane, the balloon will also be "pulled back." Much to their surprise, the physicist wins the bet.

The boy and the other passengers are in Kim's "epistemic quandary." Given that their initial expectation that the balloon would float toward the back of the plane was foiled, they are now in need of an explanation as to why the balloon acted as it did. Salmon argues that there are at least two legitimate explanations:

> First, one can tell a story about the behavior of the molecules that made up the air in the cabin, explaining how the rear wall collided with the nearby molecules when it began its forward motion, thus creating a pressure gradient from back to front of the cabin. This pressure gradient imposed an unbalanced force on the back side of the balloon, causing it to move forward with respect to the walls of the cabin. Second, one can cite an extremely general physical principle, Einstein's *principle of equivalence*, according to which an acceleration is physically equivalent to a gravitational field. Since helium-filled balloons tend to rise in the atmosphere in the earth's gravitational field, they will move forward when the airplane accelerates, reacting just as they would if a gravitational field were suddenly placed behind the rear wall. (Salmon, 183)

Salmon refers to the explanation based in molecular mechanics as explanation$_1$ and that based on the principle of equivalence as explanation$_2$. As Salmon claims, "it is my

present conviction that both of these explanations are legitimate and that each is illuminating in its own way" (ibid., 183). It would be difficult to dispute Salmon on this point. Both explanations seem to have equal claim on the explanandum event (or, the event to be explained).

Thus, there are at least two explanations that adequately account for the phenomenon of the balloon's movement. The question now is whether each is a complete and independent explanation of the explanandum event. There is not much in the way of disputing that each is a *complete* explanation. Neither explanation (as a specification of the cause of the balloon's movement) is incomplete by not referring to the principles invoked by its rival.

Is one of the explanations somehow dependent on the other? Such an argument would be equally difficult to make. The project of unifying contemporary space-time physics with atomic physics, either by bridge laws or by underlying, unifying laws, has stymied the best of physicists. So, at this time, there do not seem to be any convincing arguments which show that the laws and principles of general relativity either supervene on or reduce to the laws of atomic physics, and vice versa.

Let us return to our example: The plane accelerates, the balloon moves forward, and the physicist wins the bet. Of course, the young boy and his fellow passengers with whom he had the bet, are in an epistemic quandary with regard to the balloon's movement. An explanation from the physicist is

required in order to remove them from this quandary. But which explanation ought he give? How can the physicist give an adequate explanation of the balloon's movement when such an event admits of two apparently complete and independent explanations?

According to Kim, the physicist cannot simply choose one explanation over the other: "It is not implausible to think that failing to mention either of the overdetermining causes gives a misleading and incomplete picture of what happened, and that both causes should figure in any complete explanation of the event." (Kim, 252) Further, the physicist cannot simply give both explanations, since the puzzlement of both the boy and passengers would simply be increased. Not only would they still not know why the balloon moved as it did, but they would have the additional quandary of not knowing which explanation is correct (or, at least, how they are related).

Here, we have a real problem for Kim: The physicist, constrained by the EEP, cannot explain to the passengers why the balloon moved as it did. Clearly, this inability is not for lack of a good explanation. As Salmon claims, both explanations appear to be legitimate. But having a good explanation is not enough. According to Kim, if one appears to have multiple explanations, then one must *also* know how the explanations are related. Given that both explanations are (apparently) inescapably complete and independent, the

physicist cannot give an account of how they are related, viz. reduction, supervenience, and the like.

How is the movement of the balloon to be explained? The physicist, according to the constraints of the EEP, must say to the passengers, "I don't know." But it is obvious that he does know. Individually, there is little that is unclear or controversial about either explanation. How, then, does the physicist explain the movement of the balloon? Salmon offers an answer:

> Pragmatic considerations determine which ... [explanation] is appropriate in any given explanatory context. In the case of the friendly physicist, for example, an appeal to Einstein's equivalence principle would have been totally inappropriate; however, the [atomic] ... explanation might have been made intelligible to the boy and the other interested adults. (Salmon, 185)

In this response, the choice between the two explanations has nothing to do with the relative strength or legitimacy of either one. According to Salmon, the choice is determined by the audience to whom the explanation is given. The principle-of-equivalence explanation requires a high level of abstraction. Given this level of abstraction, along with the extreme complexity of general relativity, making such an explanation intelligible to an untutored audience would be an arduous and frustrating task for all involved. On the other hand, the physicist might have a considerably easier time getting his fellow passengers to understand an explanation based in atomic physics (at least in broad strokes), since

the level of abstraction in such an explanation would not be nearly as considerable. Further, such an explanation readily admits of visual metaphor.

For Salmon, the choice between explanations is governed by "pragmatic considerations." These considerations are focused not on the explanation itself, but on the subject or subjects to whom the explanation is given. Thus, while each explanation has legitimate claim over the motion of the balloon, one may be more "appropriate" than the other, depending on the audience to whom the explanation is given.

In what remains of this chapter, it will be shown, first, that Kim's arguments for the explanatory exclusion principle presuppose that the subject (i.e. the subject to whom the explanation is offered--the "explainee") plays a central and necessary role in explanation theory; second, an understanding of the subject's role in explanation will show that the EEP is counterproductive to the epistemic goals which Kim invokes in support of the EEP. In short, the subject is indispensable to Kim's arguments for the explanatory exclusion principle, and yet the subject renders the EEP untenable.

3. Epistemic Puzzlement and Why-Questions

Salmon indicates that each explanation is legitimate because each "provides a different kind of understanding of the same fact ... Each is illuminating in its own way" (ibid.

183). Different (satisfactory) explanations, in some cases, may be able to provide a different understanding of a particular phenomenon. In the present example, Salmon labels the molecular-based explanation a "bottom-up" explanation and the relativity-based explanation "top-down": "*Bottom-up* explanations ... appeal to the underlying micro-structure of what they endeavor to explain", while *top-down* explanations are "global ... They relate to the structure of the whole universe." (ibid. 184) Thus, each explanation generates a different perspective on the explanandum event. Each provides a satisfactory way of understanding why the phenomenon occurred the way it did.

What makes Salmon's remarks interesting is that, traditionally, theories of explanation focus on what explanations *are*. Such is the case with Hempel's covering-law model: Explanations are arguments, either inductive or deductive, where the conclusion of the argument is the event or fact to be explained (i.e. the "explanandum") and the premises (which must include a general, relevant law) are the "explanans." In such an analysis, little attention is paid to what explanations *do*, i.e. the function of the explanation.

For Salmon, the function of explanation is to provide understanding of or perspective on a fact or event. In his example, the passengers' expectations regarding the movement of the balloon are foiled. The event is anomalous for them. The function of a good explanation of the balloon's movement

is to provide the passengers with an understanding of or perspective on the event.

In considering the function of explanation, and thus invoking concepts such as "perspective" and "understanding", a factor is introduced which is largely ignored as unimportant or problematic in the more traditional models of explanation: The subject to whom the explanation is given. "Perspective" and "understanding" are relational terms, where explanation X provides an understanding of or a perspective on the object or topic of X to a subject S. As will be shown, the subject is to play a vital role in the rejection of the EEP and in a subsequent understanding of the relationship between multiple explanations.

At this point, the concepts of "perspective" and "understanding" are quite vague, but extremely important. What constitutes a perspective? What does it mean to "acquire" an understanding of an event or fact? What is the relationship between correctly explaining a phenomenon and understanding a phenomenon? In order to be useful, these concepts need clarification.

We begin with the question, what does it mean to "have" or "acquire" or "be in" a certain perspective? A beginning of an answer to this question can be found in Kim's commentary on the purpose of explanation. Recall Kim's epistemological considerations: Kim characterizes the situation in which we need an explanation as a "state of puzzlement" or an "epistemic predicament." Satisfactory

explanations remove one from this state of puzzlement, unless, according to the EEP, we have too many.

What Kim has in mind here, as he himself claims, is Sylvian Bromberger's notion of a "predicament." While there are various types of predicaments on Bromberger's account, they share at least one important common characteristic: A subject A is in a predicament with regard to a particular question Q. As Bromberger says, "'A is in a ... predicament with regard to Q' is true if and only if the question mentioned in it admits of a right answer, but that answer is beyond what [A] can conceive" (Bromberger, 36).

Therefore, this puzzlement or predicament, of which both Kim and Bromberger speak, is relative to a particular question: One is in such a predicament if one asks a question Q, where Q is a meaningful question that has a correct answer, but one cannot conceive of the satisfactory or correct answer to Q. One finds one's way out of this predicament when Q has been satisfactorily answered. Thus, it seems that, on Kim's epistemic considerations of explanation, whether or not an explanation is adequate (i.e. whether it removes one from this state of puzzlement) necessarily depends on whether the explanation satisfactorily answers the question Q, where Q is the question to which the subject's puzzlement is relative.

These considerations jibe nicely with Kim's own formulation of the epistemological version of the EEP: "No one may accept both explanations unless one has an

appropriate account of how they are related to each other." (Kim I, 257) Multiple complete and independent explanations are epistemically counterproductive, i.e. they do not remove puzzlement. In other words, multiple complete and independent explanations do not answer the subject's question. Indeed, not only do they not answer the question which expresses the subject's puzzlement, they also serve to introduce a new question: "Which explanation correctly answers my original question?"

Puzzlement is compounded since new questions or puzzlements are introduced without resolving the original questions or puzzlements. Therefore, the epistemological motivations for the explanatory exclusion principle must necessarily include considerations of the questions which express the puzzlement of the subject. These considerations must include an account of when and how such puzzlement is removed, i.e. they must include the criteria for satisfactory answers to such questions. These considerations, therefore, *are directly implied by Kim's epistemic arguments for the explanatory exclusion principle.*

Clearly, "puzzlement" is a central concept in Kim's epistemological arguments for the EEP. And its importance is paralleled only by its vagueness. Thus, it is necessary to clarify what exactly is meant by "puzzlement."

There are at least two ways of construing puzzlement--as a psychological phenomenon and as an epistemic phenomenon. In the former case, one may "feel" puzzled and such a feeling

may be relieved simply by forgetting the question that expressed the puzzlement or by taking a drug which would cause the feeling of puzzlement to dissipate, even if the question lingers. Even an answer which is false might relieve psychological puzzlement. Clearly, this is not the type of puzzlement relevant here. The specific features which characterize epistemic puzzlement can be divined from Hintikka's explication of the semantics of questions[1].

Every wh-question implies two concepts: a "presupposition" and an epistemic goal[2]. The presupposition is a description of the epistemic state of affairs necessary for the asking of a wh-question. An example of such a wh-question, asked by a subject S, might be:

(1) "Where has Smith taken Jones' car?"

For question (1), the presupposition would be the propositional attitude,

(2) S knows that Smith has taken Jones' car.

---

[1] The examples used here to explicate the epistemic features of puzzlement can be generalized through Hintikka's epistemic logic of questions and answers. Such an account can be found in "Semantics and Pragmatics for Why-Questions", in The Journal of Philosophy (1995), p. 636-657.

[2] Although Hintikka refers to this latter concept as the "desideratum," for the sake of ease and clarity, we shall use the term "epistemic goal." See ibid. p. 637.

Notice that there are two necessary conditions for the legitimacy of the question. Where A= 'Smith has Jones' car', it is necessary that (a) "S knows that A" is true (in other words, the presupposition must be true) and (b) "A" is true[3]. If (a) were false, then S would not be in a position to ask question (1). If (b) were false, then (1) would not admit of a correct answer. For instance, if Smith had not taken Jones' car, then the question, "where has Smith taken Jones' car?" cannot be given a good answer. Thus, (via conditions (a) and (b)) the presupposition of a question is a necessary condition for the asking of question Q.

Question Q, however, indicates that there is an epistemic state of affairs that subject S is lacking. This state of affairs is expressed in the epistemic goal of the question: "The [epistemic goal] of a question is a description of the state of affairs that the questioner would like to have brought about" (Hintikka, 638). Relative to (1), the epistemic goal can be expressed as such:

(3) S knows where Smith has taken Jones' car.

If question (1) is answered correctly, then (3), the epistemic goal, becomes true. The function of answers to such questions is to bring about the epistemic state of affairs expressed by the question's epistemic goal. As

---

[3] Notice that, while (a) implies (b), (b) does not imply (a). Thus, as conditions for question legitimacy (a) must be distinguished from (b).

Hintikka says, "the outcome of a conclusive answer to a question is the extra information ... of its [epistemic goal] as compared with the information conveyed by its presupposition" (ibid., 639).

While an explication of the logical features of questions and their answers is necessary to understand the general relationship between presupposition, wh-question and epistemic goal, such an explication is not necessary for present purposes. This brief sketch is enough to understand the conditions of epistemic-puzzlement and the method by which such puzzlement is resolved: For every question Q, there is implied an epistemic goal of Q and a presupposition of Q. The epistemic goal expresses the epistemic state of affairs which will be achieved once the question is satisfactorily answered. The presupposition expresses the epistemic state of the questioner at the time the question is asked. The epistemic gap between presupposition and one's epistemic goal is the puzzlement expressed by Q.

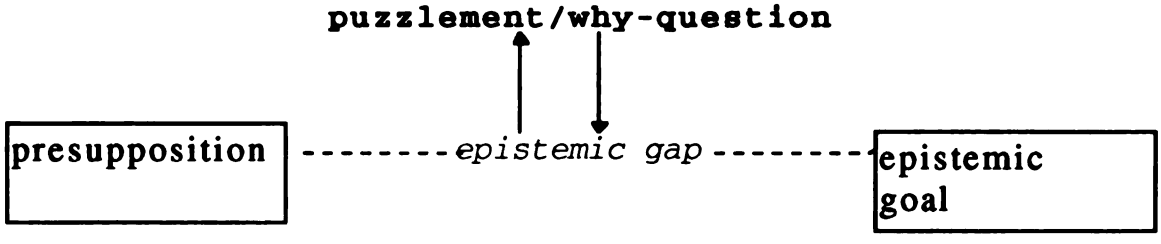

Figure 1

The state of puzzlement (of the questioner) is expressed by the question Q. The alleviation of that puzzlement is

characterized by the information gathered (i.e. a satisfactory answer to Q) which allows the questioner to move from the epistemic state of affairs expressed in the presupposition of Q to the epistemic state of affairs expressed in the epistemic goal of Q. This "epistemic movement" cannot be accomplished by forgetting Q or by taking a drug which dissipates any psychological feeling of puzzlement. These considerations of the relationship between the presupposition and the epistemic goal of Q clearly allow us to isolate this discussion from any psychological conception of puzzlement.

3. Questions and Contexts

Let us consider those criteria which, at least in part, determine whether or not an explanation is satisfactory, i.e. when the movement from the presupposition of Q to the epistemic goal of Q has been successfully achieved. Of course, that the explanation-candidate is true is a necessary condition for its being satisfactory, but is it sufficient? Will the truth of an explanation-candidate suffice to answer the relevant question effectively and eliminate puzzlement?

It is possible for an explanation to be true and still not remove puzzlement, or answer the question. Alan Garfinkel uses a famous example to illustrate this point: "When Willie Sutton was in prison, a priest who was trying to reform him asked him why he robbed banks. "Well," Sutton

replied, "that's where the money is"." (Garfinkel, 21) Sutton certainly thinks that he has answered the priest's question, and his explanation is true. However, it is doubtful that the priest's puzzlement over why Sutton robs banks is abated. It seems that they have each considered this question in different ways. Each has a different understanding of the question.

Garfinkel characterizes this difference in terms of a varying "space of alternatives." (21) Sutton presupposes a certain set of alternatives to the priest's question: "Why do you rob banks, as opposed to diners, grocery stores, private homes, etc.?" So the alternatives in this question are in contrast to banks, i.e. objects of robbing. Thus, the space of alternatives includes banks, diners, grocery stores, private homes, etc.

Of course, the priest is interested in a different set of alternatives: "Why do you rob banks, as opposed to taking up an honest, legitimate profession?" Here, the alternative is in contrast to the activity of robbing rather than the object that is robbed. So the space of alternatives includes robbing banks and alternatives to the act or occupation of robbing banks. The misunderstanding that takes place between Sutton and the priest can therefore be characterized initially as a difference in alternative spaces.

Certainly, the two alternative spaces understood by Sutton and the priest are not the only possible interpretations of the question. A wide variety of

alternative spaces are possible. For instance, one might contrast robbing banks with painting banks, building banks, demolishing banks, etc. If such a variety of alternative spaces are possible, how does one determine which alternative space is invoked in a particular question? In general, the particular alternative space of a question asked is determined by the context of the situation[4].

The response that Sutton gives to the priest's question is generally regarded as comical or humorous, since Sutton's response is not only a clear misunderstanding of the question, but it is also quite unexpected. It is unexpected to a general audience, because such an audience will more than likely have understood the alternative space implied by the priest's version of the question. This is because a general audience would recognize certain important features of the context.

For instance, they would recognize the questioner as a priest (by clues such as a collar or a bible) and understand that the priest is interested in reforming Sutton, or at least in getting Sutton to repent his sins. Most observers of this situation, knowing what they know about priests, will not make the same misinterpretation that Sutton makes. This makes his misinterpretation both unexpected and comical. Clearly, the priest is not interested (as Sutton interprets)

---

[4] Note that the present use of the term "context" is non-technical. The "context of the situation" is simply the salient features of the situation in which the question is asked--features the might provide clues as to the intended meaning of the question.

in the advantages of robbing banks instead of grocery stores, private residences, etc.

Ascertaining certain features of the context of the situation, namely that a priest is asking the question and priests have certain aims and roles to play in such situations, allows a general audience to understand the alternative space of the priest's version of the question. Sutton either ignores or is ignorant of such features of the context, which allows him to misinterpret the priest's question.

If those contextual features were changed, so would the interpretation of the question. For example, if, instead of a priest, the questioner is a friend of Sutton's and a thief, Sutton's interpretation of the question (that is, Sutton's alternative space) would be quite reasonable. Thus, changing certain features of the context in which the question is asked alters the conditions by which the question is interpreted. In this latter context, Sutton's interpretation of and answer to the question, "why do you rob banks?" would be reasonable and, probably, appropriate.

So far, the relation between contexts, questions and alternative spaces has been painted in rather broad strokes. It is, at this point, necessary to supply some detail[5]. The Sutton example illustrates that it is possible (though, as will be shown, not essential) to ask questions which express

---

[5] A good deal of this detail is drawn from Bas van Fraassen's work. See specifically "A Theory of Why Questions" in The Scientific Image.

different alternative spaces with the same utterance. Thus, such an example allows us to distinguish between an interrogative utterance and a question (or interrogative proposition) where the question is characterized (in part) by its alternative space.

The priest's utterance, "Why do you rob banks?" is taken to express two different questions and one can distinguish between the questions by understanding their different alternative spaces. Further, coming to understand the question expressed in an interrogative sentence can often be accomplished by understanding the context in which the sentence is uttered.

There is still a good deal of detail which remains unexamined regarding the relationship between interrogative sentence, question and context[6], but the detail given thus far is enough for present purposes. The upshot of these considerations is this: According to Kim, explanations remove a subject from a state of puzzlement. That is their epistemic function. Puzzlement, in order to be philosophically respectable, must first be expressed in a why-question Q. More specifically, puzzlement is the epistemic gap between the presupposition of Q and the epistemic goal of Q. To relieve a subject of his or her puzzlement is to bridge the gap between presupposition and

---

[6] Some of these details will be addressed in chapter III.

epistemic goal. This is accomplished by successfully answering the subject's question.

But in order to answer such a question and thereby alleviate epistemic puzzlement, one must understand the question that is asked. However, the Sutton example illustrates that in order to understand (and thus successfully answer) a particular question, one needs to understand (at least) the alternative space of the question. If such an understanding is not achieved, then one runs the risk of misunderstanding the question and giving an unsatisfactory answer. Such is the case with Willie Sutton's answer to the priest's question. Clearly, Sutton's answer does not relieve the priest of his epistemic puzzlement regarding Sutton's chosen profession, even though his explanation is true.

But what, in general, are the conditions by which a question is successfully answered? Recall that, in the priest's understanding of the interrogative, "Why do you rob banks?" the robbing of banks is set up in opposition to earning an honest wage, or living an honest life, i.e. "Why do you rob banks, *as opposed to* earning an honest wage?" In other words, the robbing of banks (what van Fraassen calls the ***topic*** of the question, "Why do you rob banks?") is a member of the space of alternatives, amongst other possible alternatives; in this case, earning an honest wage. Broadly speaking, a successful answer to a question will select the topic of the question from the space of alternatives.

For example, Sutton's answer, "That's where the money is" is a good answer to the question "Why do you rob banks?" where the alternative space includes {robbing banks, robbing grocery stores, robbing trains, robbing private residences, etc.}, because it selects "robbing banks" from the other alternatives. The fact that banks have the most money is why one robs banks and *not* grocery stores, trains, private residences, etc.

So, in general, a good answer selects the topic from the range of alternatives. In other words, an answer shows why the topic is true and the alternatives are false. This notion of a successful answer fits quite nicely with Hintikka's idea of epistemic movement from presupposition, "I know that Willie Sutton robs banks," to epistemic goal, "I know why Willie Sutton robs banks", as opposed to grocery stores, trains, etc.

Again, these considerations of the nature of puzzlement, why-questions, context, and alternative spaces follow from the epistemic foundations of the explanatory exclusion principle. The epistemic aim of explanation is to remove the subject from their state of puzzlement. The conditions of such a removal are made explicit by the above considerations. If it turns out that the EEP runs counter to this explication of puzzlement removal, then it also runs counter to the epistemic aims of explanation which are supposed to support it. Such a situation would do more than simply show that the EEP is problematic, as the examples given by Scriven, Fodor

and Salmon illustrate. Such a situation would render the EEP untenable, since it would undercut its own justificatory principles.

4. The Epistemic Aims of Explanation

If the above considerations are correct, then (as will be shown) the EEP is counterproductive to the very epistemic aims of explanation which are supposed to serve as its justification. Thus, the EEP is untenable and therefore not a reasonable condition on explanation in general. The final section of this chapter is dedicated to the support of this claim.

Before such an argument can be made, there is an important addendum to the considerations regarding the relationship between interrogative sentence and question. The Sutton example illustrates how one can understand the utterance of an interrogative sentence and yet misunderstand the question that is being asked. In such a case, the same interrogative sentence is capable of indicating multiple questions, viz. multiple alternative spaces. However, nothing crucial hangs on the idea that Sutton and the priest use the same utterance, or interrogative sentence, to express different questions. Certainly, the priest's interrogative sentence might have been phrased differently, so as to allow Sutton to understand the question he had in mind. The general fact that different questions can be asked through

the same interrogative utterance allows us to distinguish between questions and interrogative sentences. It is not essential that in every example, the same interrogative must be given to express different questions.

How, then, do the above consideration serve as a response to Kim and the EEP? Let us take another example. Smith and Jones see their friend Roy walking down the street. Smith is aware that Roy has a neurological condition that, when untreated, causes his right arm to spasm. Jones is not so aware. Suddenly, Roy begins waving his arm wildly. Smith utters the interrogative, "Why is Roy's arm flailing about?" Jones, unaware of Roy's condition, assumes Roy is hailing a nearby cab and utters, "Why is Roy hailing a cab?"

Given that Smith is aware of Roy's neurological condition, it is reasonable to assume the alternative space {Roy's hand is waving, Roy's hand remains at his side}. Jones, on the other hand, knows that Roy is on a fixed income and cannot afford extravagances. Thus, it is reasonable to assume the alternative space {Roy is hailing a cab, Roy takes a bus, Roy walks to his destination, Roy hitches a ride}. After all, Roy is on a fixed income and can ill afford to spend money on a cab. Why, then, is he hailing a cab, as opposed to simply taking the bus or walking or hitching?

Let us suppose two possible scenarios as outcomes. Scenario1: Roy sits down on a park bench, his arm still waving. He takes his medicine and relaxes. After a few minutes, the movement of his arm subsides. Scenario2: A cab

pulls up to the curb near Roy. Roy stops waving his hand, gets in, instructs the driver and the cab departs.

It is clear that in scenario1, Roy is not hailing a cab at all. Thus, Jones' question, "Why is Roy hailing a cab?" is misguided. In fact, characterizing the movement of Roy's arm in an intentional matter would be inappropriate, since it is obviously not an intentional action. As his subsequent actions suggests, his behavior is a result of his neurological condition and can be appropriately characterized as a spasm, rather than a purposive action. On the other hand, Smith's question is entirely appropriate. The answer, "Roy's arm-movement is a result of his neurological condition"[7] picks out the topic of Smith's question from its alternative. These considerations appear to agree with Kim's analysis. The neurological idiom allows Smith to ask a legitimate question and receive a good answer, whereas the intentional idiom has failed Jones. His characterization of Roy's behavior as intentional, implied by his question, is inappropriate.

Scenario2, however, has radically different consequences for Smith and Jones' questions. It is clear that in scenario2, Roy is indeed hailing a cab. On Hintikka's analysis, Jones' question is appropriate, since it meets both conditions for question legitimacy: Where the question is, "Why is Roy hailing a cab?" it is necessary that "Roy is

---

[7] Certainly, a more detailed answer, regarding how the neurological disorder caused the movement, would be possible, but the given answer is sufficient for present purposes.

hailing a cab" is true and "I [Jones] know that Roy is hailing a cab" is true. The context of scenario2 determines that Jones' question is legitimate, whereas the context of scenario1 determines that it is not. In this way, whether or not the topic of a particular question is true is determined by the context of the situation.

Since, in scenario2, Jones' question is legitimate (i.e. it meets both of Hintikka's conditions) it must admit of a correct answer or explanation. In other words, there is an answer which makes the topic of the alternative space true and the alternatives to the topic false. Notice, however, that since the topic (and thus the question) and the other members of the alternative space are characterized intentionally, the answer which selects the topic from the alternative space must make reference to Roy's intentional states. For example, one might offer the answer "Roy believed he was late for work and that a cab would be the fastest way to get there." Such an answer (if true) would select the topic from the space of alternatives. In other words, it would explain why he took a cab and did not take the bus, hitch a ride or walk.

But why must an answer to Jones' question invoke Roy's intentional states? Wouldn't an answer such as "Roy is late for work. The fastest way to work is by cab," be a good answer? Why do Roy's intentional states have to enter the picture at all?

Here, the unique character of propositional attitudes plays a crucial role. There are a number of possible situations (e.g. Roy's watch is fast, Roy forgot that daylight savings time ended, etc.) in which "Roy is late for work" or "the fastest way to work is by cab" are false, and yet "Roy believes that he is late for work" or "Roy believes that the fastest way to work is by cab" are true. Whether or not Roy *is* late for work is irrelevant to Roy's hailing a cab. It might very well be false. What is relevant is that Roy *believes* he is late for work, regardless of whether or not he is, in fact, late. Thus, in scenario2, an answer to Jones' question, if it is to select the topic from the alternative space, must involve reference to Roy's intentional states--his beliefs and desires.

Here, we come to the heart of the problem: Jones' puzzlement is expressed in a particular why-question. The alleviation of Jones' puzzlement (the epistemic goal of explanation, according to Kim) requires an answer to her question which necessarily includes reference to Roy's intentional states, because such an answer will pick out "Roy is hailing a cab" from amongst other possible intentional actions in the alternative space. However, in principle, the explanatory exclusion principle rules out the possibility of invoking such intentional states in explanations, unless the states posited in the intentional idiom are somehow reducible to (i.e. systematically identifiable with) the states of the neurophysiological idiom. But arguments such as multiple

realizability preclude any such systematic reduction. Therefore, according to the EEP, such intentional states cannot enter into explanations of behavior.

But this simply means that, in principle, Jones' puzzlement, as expressed in her why-question, cannot be alleviated, since a necessary condition for answering her question and thus alleviating her puzzlement must include reference to Roy's intentional states. But recall that abating puzzlement is the epistemic goal of explanation, as stated by Kim himself:

> I take it that *explaining* is an epistemological activity ... To be in need of an explanation is to be in an epistemically incomplete and imperfect state, and to gain an explanation is to improve one's epistemic situation; it represents an epistemic gain. (Kim I, 255-6)

Kim invokes this argument as a justification for the EEP. However, if we are constrained by the EEP, then it must be the case that any question the answer to which must include reference to intentional states, such as Jones' question, cannot be answered. The puzzlement cannot be abated. The "epistemic gains" cannot be achieved.

The problem for Kim is compounded if we consider Fodor's arguments. The EEP can be applied equally well to all the non-physical sciences. Thus, any why-question the answer to which must refer to the systematically irreducible states of any non-physical sciences cannot be answered. The EEP must rule out, in principle, the possibility that any such

question can be answered and therefore the puzzlement expressed in those questions cannot be abated.

Clearly, if the alleviation of puzzlement is the epistemic goal of explanation, then rather than being supported by this goal, the explanatory exclusion principle is counterproductive to it--to an extreme degree. Therefore, the EEP is not, as Kim believes, a reasonable constraint on explanation in general. On the contrary, it is counterproductive to the very goals of explanation that Kim sets out in support of the EEP.

One might offer an objection to this argument: Doesn't there exist a kind of asymmetry between the intentional and neurophysiological idioms? In all contexts (scenarios), it is possible to ask a question for which a neurophysiological answer is appropriate. Even in scenario2, in which Roy's behavior can legitimately be characterized intentionally, one can also ask a question whose answer must involve reference to Roy's neurophysiological states.

The same cannot be said for the intentional idiom. There are cases, such as scenario1, in which an intentional characterization of behavior is illegitimate. In short, for every behavioral event, there is a possible neurophysiological explanation. But it is not the case that for every behavioral event there is a possible intentional explanation. Jones' question in scenario1 is just such a case. Doesn't this suggest a certain explanatory priority of the physical over the intentional? Don't we get better

explanations from the neurophysiological idiom because the idiom is not subject to explanatory or descriptive failure?

Certainly, there is an asymmetry between the intentional and the neuro-physical. While this asymmetry does imply a difference in explanatory scope, it does not imply a general priority of the neurophysiological idiom over the intentional. In cases such as scenario1, the neurophysiological idiom is superior to the intentional in that Roy's behavior admits of physical description, but not to intentional description. But this does not imply a superiority in general.

In cases such as scenario2, the neurophysiological idiom can give a description of Roy's behavior and therefore answer certain questions about it. But it does not follow that it can answer *all* possible questions about Roy's behavior. Jones' inquiry is an example of such a question. Therefore, while the neurophysiological idiom is superior to the intentional in that the former does not admit (in general) of explanatory failure, it does not follow that neurophysical explanations are superior to intentional explanations in principle.

Let us consider a final objection: In Smith's question, the topic is "Roy's hand is flailing" or "Roy's hand is moving," whereas in Jones' question, the topic is "Roy is hailing a cab." Isn't there a sense in which they are not explaining the same thing? There is a difference between explaining the act of cab-hailing and the act of arm-

flailing. Can't we simply say that they are explaining different events?

Kim confronts a similar objection to his arguments for the explanatory exclusion principle. Recall from chapter I the example of Smith's climbing a ladder to retrieve her hat from the roof of the garage. There are two possible explanations of this behavior, one from the intentional idiom, the other from the neurophysiological idiom. The difficulties of the explanatory exclusion principle arise when both explanations are said to give complete and independent explanations of Smith's hat-retrieval behavior.

But one might object to Kim that the problems of having multiple explanations do not arise, since there is more than one event to be explained. In the neurophysical explanation, the event to be explained (the explanandum) is expressed in non-intentional, "mere physical movement" terms. In the intentional explanation, the event is explained in intentional, purposive terms. The problems of explanatory exclusion arise only when there are multiple explanations of a single explanandum event. Such is not the case here.

Kim responds that, while there are multiple descriptions of Smith's behavior, there is a sense in which all descriptions are "about" the same event: "Although the two explanandum statements are not equivalent or synonymous, there is an evident sense in which they 'describe' one and the same event, the same concrete happening." (Kim I, 242)

I believe this is a reasonable response to this type of objection. Regarding Roy's arm-movement, while there may be many different ways to describe Roy's behavior, and thus many different question-topics about his behavior, they all have the same referent; namely, Roy's bodily activity at that time. Thus, while different questions are being answered and different explanations are given, they all explain the same event, given that the topic of each question has a referent identical to the topic of the other. So answers to both Smith and Jones' questions will be different, but they will be explanations of the same event.

The explanatory exclusion principle is not, as Kim claims, a reasonable constraint on explanation in general. On the contrary, it is counterproductive to the epistemic goals of explanation. The problems with the explanatory exclusion principle illustrate that we must develop an analysis of mental causation which makes sense of and makes room for the causal efficacy of irreducible mental states in the production of bodily behavior, given that the physical idiom is (in principle) capable of providing exhaustive causal analyses of any behavioral event, intentional or otherwise. Further, we now have the key which will allow for such an analysis of mental causation: A pragmatic conception of explanation, i.e. a conception of explanation in which the subject to whom the explanation is offered plays a central role. What is required is a fuller understanding of the

pragmatics of explanation, such that a positive analysis of mental causation can be given.

Chapter III: The Pragmatic Conception of Explanation

1. A Preliminary Solution

The above considerations make clear that the explanatory exclusion principle is too strong a condition on explanation in general. Thus, we need not rule out, in principle, the possibility of the intentional idiom providing complete and independent explanations of bodily behavior. But the original problem remains: Physicalism entails that for every behavioral event, one can give an exhaustive causal analysis within the physical idiom. If mental states are systematically irreducible to physical states, then their properties (including causal properties) cannot be assimilated into this causal analysis. Thus, while we need not rule out intentional explanations in principle, it is not yet clear how we can make room for such causal explanations, relative to the exhaustive causal analyses that the physical idiom is capable of providing.

At best, it is not clear what causal work remains for mental states and properties if all physical behavior has a sufficient physical cause. At worst, attributing causal efficacy to irreducible mental states runs the risk of encountering the metaphysical and epistemological difficulties that (supposedly) motivate the explanatory exclusion principle; the most important of which is causal overdetermination. While we need not heed the explanatory

exclusion principle and exclude intentional explanations in principle, giving multiple complete and independent explanations still smacks of systematic overdetermination[1].

Thus, a rejection of the explanatory exclusion principle does not automatically relieve us of the original difficulties which plague the idea that mental states are both irreducible and causally efficacious. Our question remains: How do we make sense of the causal work done by irreducible mental states within a physicalist conception of the world?

The answer to this question lies within those pragmatic considerations of explanation briefly explicated in chapter II. The role of the mental in causing behavior can be understood through an analysis of the role of the mental in causal explanations of behavior. Within such an analysis, our question can be rephrased: How do we make sense of and make room for causal explanations provided by the intentional idiom, given that, for any behavioral event, a complete causal explanation can be given in the physical idiom?

The goal is to allow for the fact that the physical sciences are capable of giving complete explanations of any behavioral event, while maintaining the importance of irreducible, intentional explanations. This is exactly what the pragmatic considerations of explanation allow us to do.

---

[1] Systematic overdetermination, recall, is a causal overdetermination which is set up and maintained by our theories of the world, as opposed to an accidental or anomalous overdetermination. See chapter I, p. 13.

Physicalism, again, implies that any behavioral event is subject to an exhaustive causal analysis in the languages of the physical sciences. Therefore, any behavioral event can be given a complete physical explanation. On the pragmatic conception of explanation, such a claim can be understood to mean that, first, for any behavioral event, there is a why-question regarding that event (i.e. a why-question whose topic is the event) which requires an answer from within the physical idiom; second, any such appropriately-phrased why-question which requires an answer from the physical language can be answered. Such are the implications of physicalism for the pragmatics of explanation.

However, what is not implied by physicalism is that any why-question *in general* can be given a physical explanation. This is evident in the arm-waving example discussed in chapter II. The question, "Why is Roy hailing a cab?" characterizes Roy's behavior as intentional and requires an explanation which refers to Roy's intentional states. A physicalist answer can be given to explain Roy's behavior, but such an answer would be inappropriate, i.e. it would not satisfactorily answer the why-question and relieve the subject of his/her epistemic puzzlement. Roy's mental states are able to do causal work that his physical states cannot in that his mental states can provide causal explanations that his physical states cannot.

Notice, this does not jeopardize the implications of physicalism. It remains true that, for any behavioral event, a physical explanation can be given. In other words, for any behavioral event, a why-question requiring an answer couched within the physical idiom, can be asked and (in principle) can be answered. But on a pragmatic conception of explanation, this does not imply that the physical idiom can provide an answer to any why-question regarding a behavioral event. Mental states do causal work by providing answers to why-questions that cannot be answered in the physical idiom.

This analysis of the causal role of mental states in the production of behavior is, admittedly, brief. However, it is clear that the pragmatic conception of explanation plays an indispensable role in this analysis. Thus, in order to understand and justify this analysis of mental causation, we must fill in some of the details of such a theory of explanation. The remainder of this chapter is dedicated to this purpose. A more complete understanding of the pragmatic conception of explanation will lend our analysis of mental causation both clarity and credence.

2. Understanding Questions: Alternative Space, Topic and Relevance Relation

Recall Kim's claim that the epistemic function of explanation is to remove a subject from his/her state of puzzlement. The subject expresses this puzzlement in the form of a why-question. As such, puzzlement can be

understood as an "epistemic gap" between the presupposition and the epistemic goal of the question[2].

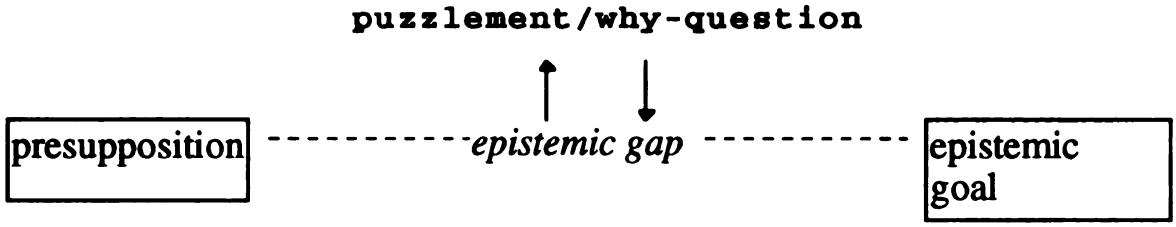

Figure 1

To answer the why-question is to allow the subject to move from presupposition to the epistemic goal. Such answers, which remove the subject from a state of puzzlement, are explanations.

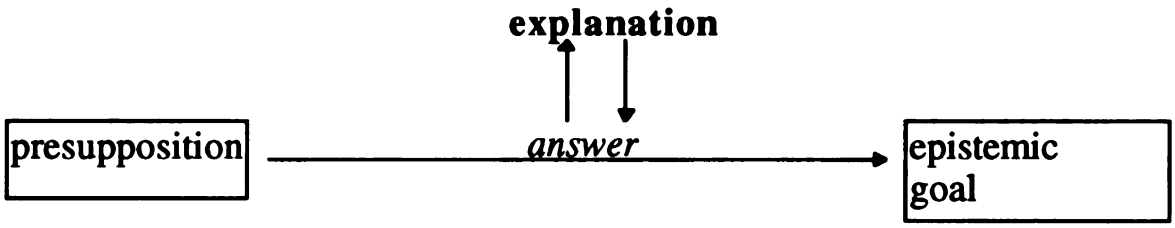

Figure 2

Therefore, the conditions by which an explanation is deemed adequate are determined by the conditions by which a why-question is successfully answered, i.e. by which the move

[2] Recall that the "presupposition" is the epistemic state of affairs necessary for the asking of a particular why question. Where the question is "Why is it the case that X?", "I know that X" is the presupposition. The epistemic state expressed by the "epistemic goal" includes "I know why X" as well as "I know that X." Again, Hintikka labels this latter concept the "desideratum" of the question. See chapter II, p. 8.

from presupposition to epistemic goal is made. An examination of such conditions was begun in chapter II. In order to understand the pragmatic theory of explanation more fully, a more detailed analysis is in order[3].

Recall the distinction between interrogative sentence and question. The subject desiring an explanation of a particular phenomenon or event utters an interrogative sentence. In order to alleviate the subject's puzzlement, the explainer must ascertain the specific question asked, so that a satisfactory answer can be given. Willie Sutton's answer, "That's where the money is," demonstrates a particular understanding of the priest's interrogative sentence, but a misunderstanding of the intended question. The present task is to enumerate the features of why-questions which must be understood, such that an appropriate answer can be given.

One such feature is the *alternative space*[4]. In order to understand a why-question, one must understand the alternative space of the question. Recall that Garfinkel characterizes the Sutton example as a mistake in alternative space. Sutton's answer selects "robbing banks" from the alternative space consisting of {robbing banks, robbing

---

[3] The analysis given will be based primarily on van Fraassen's theory of why-questions from The Scientific Image, 1980.

[4] Garfinkel's concept of the alternative space is roughly equivalent to van Fraassen's idea of the contrast class. For present purposes, I will take these terms as synonymous.

private residences, robbing grocery stores, etc.}. The priest's question, however, requires Sutton's answer to select "robbing banks" from the alternative space consisting of {robbing banks, leading an honest life with an honest job}.

However, the interrogative sentence "Why do you rob banks?" can express either question with either alternative space. Hearing the interrogative sentence uttered is not enough. One must be able to grasp (at least) the alternative space of the question which the interrogative sentence expresses. Again, we are (ceteris paribus) able to ascertain the intended question via the context of the situation[5].

In the context of the Sutton example, one notices that a priest asks the question. There are certain facts that one knows about priests; namely, that they are interested in reforming sinners and not (one would hope) in learning the best places to rob. Thus, Sutton's interpretation of the question is clearly inappropriate, given the clues provided by the situation.

To make the point clear, let us take a second example. A young girl is fascinated by a steam-engine train that rests at a train station. The wheels of the engine begin to churn

---

[5] As is the case in Chapter II, the term "context" is not used in any technical sense. Here, the term refers to the situation in which the question is asked. Clues provided by the situation allow an explainer to ascertain the question expressed by the interrogative sentence.

and the train lurches forward. The girl, in awe, asks the conductor, "Why is the train moving?" How should the conductor understand the girl's question? The context of the situation includes the fact that the inquirer is an inquisitive young girl and that she is fascinated by the movement of the train. From these facts, there may be several reasonable alternative spaces that the conductor can infer. For instance, one might assume that the girl is interested in how the train is able to move at all. Thus, a reasonable interpretation of the alternative space might be, {the train moves, the train is incapable of motion, etc.}.

Before the conductor has a chance to answer the girl, an angry young man, carrying a suitcase, hurries up to the conductor with his pocket-watch extended and demands, "Why is the train moving?" Here, the conductor is presented with the same interrogative sentence, but with a radically different situation. There are several salient features of this second scenario: The young man is carrying a suitcase, indicating he is a traveler and supposed to be on the train. He is angry, so it is doubtful he is merely indulging an interest in the mechanics of the train. Finally, he shows the conductor his pocket-watch. A reasonable alternative space might be, {the train is moving, the train remains at rest until the proper departure time}.

There may be several reasonable interpretations for each situation, and thus several different possible alternative spaces. But it is clear that the different situations in

which the utterances are made require radically different interpretations of the intended question. This difference in interpretation can be characterized, in part, by a difference in alternative space. Thus, the alternative space must be understood (via clues provided by the situation in which the question is asked) such that an appropriate answer to the question can be given.

While the alternative spaces of the questions of the young girl and the angry man are markedly different, their respective questions share at least one common feature: Both questions are about the movement of the train. In both cases, it is the motion of the train that is in need of explanation. The difference in the alternative space of each question shows that the conditions by which the motion of the train is explained will be considerably different for each question. However, there is still a basic, intuitive sense in which the movement of the train is the subject matter of both questions.

Generally speaking, where "why is it the case that X?" is the question, X is the subject matter of the explanation. In a broad sense, X is the event to be explained. As such, we are able to consider X in two different ways. First, in order for a subject to ask "Why is it the case that X?", that subject must know that X is the case. Hintikka labels knowing X (or, more specifically, "I know X") as the *presupposition* of the question. Recall that puzzlement is the epistemic gap between a presupposition and one's

epistemic goal. In order to be in a state of puzzlement, i.e. in order to ask a why-question, a certain epistemic state of affairs must obtain. That epistemic state of affairs is expressed by the presupposition. Thus, "I know that X", insofar as it is a necessary epistemic condition for the asking of "Why is it the case that X?", is the *presupposition* of the question.

However, "I know X", as the presupposition, is not itself a feature of the why-question. Since it is the epistemic state of affairs necessary for the *asking* of the question, the presupposition is antecedent to the asking of the question. On the other hand, there is a sense in which X itself *is* a feature of the why-question. Given the question, "Why is it the case that X?", a successful answer must demonstrate why X is the case, as opposed to the other possible alternatives.

Each why-question has an alternative space. The members of this alternative space include not only possible alternatives to X, but X as well. A good answer to "Why is it the case that X?" will "select" or "favor" X over the other members of the alternative space. In other words, a good answer will show why X is true, as opposed to the other alternatives (or at least show that X is more probable than its alternatives). As such, X is the *topic* of the why-question, "Why is it the case that X?"

The alternative space is simply a set of possible states of affairs or events. The topic is the actual state of

affairs, amongst a number of possibilities. A satisfactory answer to "Why is it the case that X?" must demonstrate why X is true, instead of other possible alternatives. "Why do you rob banks, as opposed to grocery stores, trains or private residences?", "Why do you rob banks, as opposed to living an honest life with an honest job?", "Why is the train moving, as opposed to remaining motionless?". A good answer to any of these questions, must select the topic from its alternatives (i.e. make the topic true), or favor the topic over its alternatives (i.e. make the topic more probable than its alternatives)[6].

Thus, in the pragmatic theory of explanation, X serves two different functions: First, X is contained in the *presupposition* of the question, insofar as knowing X is a necessary epistemic condition for the asking of "Why is it the case that X?"; second, X is the *topic* of the question, insofar as a satisfactory answer must select X from the alternative space of the question. Thus, as the topic of the question, X is a defining feature of the question and, as such, helps determine the conditions for a satisfactory answer. In short, a successful answer must select the topic from the alternative space.

---

[6] The general conditions by which an answer favors or selects the topic from the alternative space is not a central concern. However, a brief treatment will be given in section II of this chapter. For a thorough treatment of these issues, see van Fraassen, p. 146.

Finally, the conditions by which an answer is determined to be relevant must be enumerated. Such conditions are contained in what is generally called the *relevance relation* of the question. An answer-candidate must be relevant to the question at least in the sense that it must pick out the topic from the alternative space. For instance, the sentence, "The sky is blue," is true, but it is irrelevant to the question "Why is the train moving?" since it does not address the topic of the question, viz. picking out the topic from the alternative space. Thus, at a minimum, an answer is relevant if it addresses the topic of the question. But a relevant answer must do more than simply address the topic.

Recall the example of Roy's hand-waving behavior. Smith, aware that Roy has a neurological condition which, at times, causes certain muscles to seizure, asks "Why is Roy's arm waving about?". A relevant answer must select the topic from the alternative space. However, there may be other questions containing the topic, "Roy's arm is waving about." For instance, in Jones' question, "Why is Roy hailing a cab?", the topic, "Roy is hailing a cab," is equivalent to "Roy's arm is waving about" in the sense that both topics are the same behavioral event, even if they describe the event differently.

Thus, if the requirement of relevance is limited to selecting the topic from the alternative space (i.e. making the topic true), then an answer to Smith's question will also count as an answer to Jones' question, since both questions

maintain the same topic. But, as shown in chapter II, a neuro-physiological answer to Jones' question, "Why is Roy hailing a cab?", is not relevant. Therefore, a relevant answer must do more than simply make the topic true or highly probable.

First, an answer must address not only the topic, but the other members of the alternative space as well. Recall that the members of an alternative space can be enumerated by making them explicit in the why-question: "Why is it the case that X, as opposed to Y, Z, etc.?". A relevant answer must not only make the topic true or highly probable, *it must also make the other members of the alternative space false or less probable*. Let us say, a relevant answer must "select" or "favor" the topic *over* the other alternatives. In this way, an answer must address not only the topic, but all the members of the alternative space.

Second, the relevance relation does not simply require an answer-candidate to select the topic from the alternative space. The relevance relation *determines the specific conditions* by which an answer selects the topic from the alternative space. As van Fraassen says, the relevance relation "determines what shall count as a possible explanatory factor." (143)

To clarify the point, let us return to Roy's arm-waving behavior. The topic of both Jones' question, "Why is Roy hailing a cab", and Smith's question, "Why is Roy's arm waving about?" address the same event: Roy's arm motion.

However, each question differs with respect to both alternative space and relevance relation. An answer to Jones' question must not only select the topic from the alternative space, it must do so by addressing the intentions and purposes of Roy's actions. As such, the question presupposes that Roy's intentional states are causally relevant to his bodily behavior and demands that an answer-candidate capture the intentional character of his behavior. Therefore, the relevance relation of Jones' question can be labeled "causal-intentional." A relevant answer to Jones' question must select the topic from the alternative space by addressing the intentional states which brought about Roy's cab-hailing behavior.

On the other hand, Smith's question is in regards to the neuro-physiological cause of Roy's arm-waving behavior. As such, a relevant answer must select the topic of Smith's question from its alternative space by addressing the neuro-physiological cause of Roy's arm-waving. Thus, we can characterize the relevance relation of Smith's question as "causal-mechanical."

In this example, Smith's question and Jones' question have identical topics, in the sense that both topics describe the same behavioral event. However, they diverge with respect to both alternative space and relevance relation. Consequently, answers to each question must pick out the same topic, but from different alternative spaces and according to

different relevance conditions (i.e. different conditions by which an answer is judged relevant).

What in general constitutes a relevance relation is a difficult question. Indeed, such a question cuts to the core of much of the contemporary controversy surrounding theories of explanation. Establishing conditions which determine when and how an explanans is explanatorily relevant to its explanandum is one of the central and most debated topics in the current literature. In principle, there may be a number of different ways for one state of affairs to be explanatorily responsible for another.

Nevertheless, whatever else constitutes a relevance relation, the various types of causal relations (e.g. causal-intentional, causal-mechanical, etc.) must certainly be included in their number. This is the stance Michael Scriven takes in "Causation as Explanation": "Causation is the relation between explanatory factors ... and what they explain." (Scriven, 11) Often, when we seek to know why one particular state of affairs obtains, as opposed to other possible states of affairs, we are looking for an antecedent event which brought about or caused the present condition. If I were to ask the question, "Why is my window broken?", I presuppose not only that the window is broken, but that there is some antecedent event which is causally responsible for the state of my window. This antecedent event explains my window's being broken in that it is causally responsible for the present state of affairs.

I wish to leave open the possibility of other non-causal types of relevance relations. It is a matter of some controversy whether there are such non-causal relevance relations[7], but this issue need not be a concern for present purposes. Since an explication of the pragmatic theory of explanation (including the relevance relation) is important insofar as it allows us to understand mental causation, all that is required is that, whatever else is counted as a relevance relation, the "causal-mechanical" and "causal-intentional" relations are to be included in their rank.

Let us sum up: In order to understand a question expressed by an interrogative sentence, it is necessary to ascertain the alternative space, the topic and the relevance relation of the question. Van Fraassen expresses this idea in the claim that why-questions can be identified with the ordered triple <Pk, X, R>, where Pk is the topic, X is the alternative space (or, in his terms, the "contrast class") and R is the relevance relation. An answer to a question must bear relation R to <Pk, X>. In other words, an answer must select or favor the topic from the alternative space, according to the conditions specified by the relevance relation. These are the features of why-questions that must be understood if an answer or explanation is to be given successfully.

---

[7] Cf. Chapter II, p. 6. The movement of the boy's balloon can be given an explanation based on the principle of equivalence regarding acceleration and gravitational fields. It is not clear whether this explanation is causal.

3. Explanatory Context: On What We Need To Know in Order To Be Puzzled

As Kim claims, the epistemic goal of explanation is to relieve a subject from a state of puzzlement, where such puzzlement is expressed in the form of a why-question. However, as was discussed in chapter II, there are ways of relieving puzzlement which do not answer a subject's question. A good television show or a drug is enough to relieve a subject's puzzlement by making him forget both the question and the feeling of perplexity or puzzlement. But neither television shows nor drugs count as good explanations. On the other hand, puzzlement can be relieved by answering the question which expresses the subject's puzzlement. Such answers are explanations and it is this method of puzzlement-removal which is our central concern.

Hintikka's analysis of wh-questions allows us to discern this method from the former, bogus methods of puzzlement-removal: For every such question, there is a presupposition and an epistemic goal. The presupposition is the epistemic condition necessary for the asking of a particular question (i.e. the conditions necessary for being in such a state of puzzlement). For instance, in order to ask, "Why did Mike Tyson throw the fight?", it is necessary to know that Mike Tyson threw the fight (as well as what it means to "throw" a fight). The explanatory goals of a question are the epistemic state of affairs which obtain when a question has been successfully answered. Thus, the epistemic goal of "Why

did Mike Tyson throw the fight?" would be to know why Mike Tyson threw the fight.

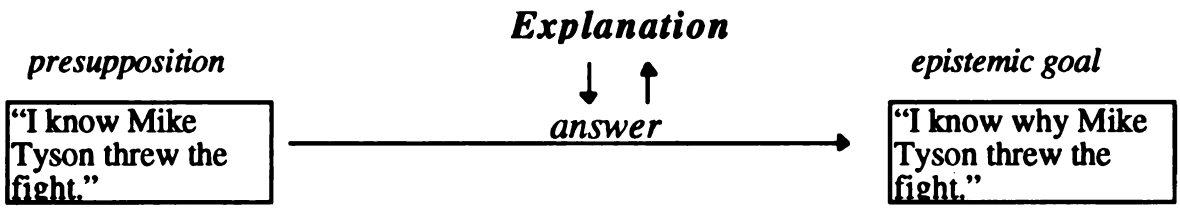

Figure 3

Notice, however, that on this analysis of puzzlement, there is much that one needs to know in order to be puzzled. To ask a question, "why is it the case that X?", one needs to know "It is the case that X." One needs to know something about what she does not know.

Recall the example of Roy's hand-waving behavior. Jones asks, "Why is Roy hailing a cab?" Jones is puzzled about Roy's behavior and expresses her puzzlement in the form of a why-question. However, in order to be puzzled at all, there is a good deal that Jones must understand about Roy's behavior. In order to ask such a question, Jones must know that Roy is, indeed, hailing a cab. Generally speaking, in order to ask such a question, the epistemic state of affairs expressed in the presupposition of the question must obtain for Jones.

Further, in addition to the epistemic conditions expressed in the presupposition, Jones must also know what it means to hail a cab. Further, hailing a cab is an

intentional action. In order to ask such a question (and understand the response), Jones must be able to characterize (or identify) Roy's behavior as intentional. Notice also that the alleviation of puzzlement occurs when Jones' question has been successfully answered. But, in order to be successful, an answer must pick out the topic from the alternative space according to the conditions specified in the relevance relation. In this case, the relevance relation is "causal-intentional." Therefore, not only are there epistemic conditions necessary for Jones' question, there are also theoretical conditions. Not only must Jones know that Roy is, in fact, hailing a cab, Jones must also have an ontology which includes cabs and acts such as "hailing," and know that Jones has intentional states which are capable of causing purposive action.

The upshot of these considerations is that epistemic puzzlement cannot occur in a theoretical vacuum. On the present conception of why-questions and the conditions by which such questions are answered, the asking of why-questions necessarily makes both theoretical and epistemic presuppositions. Such a claim is similar to the now common idea that experience, in general, is theory laden. Puzzlement, question-asking and question-answering are also

theory-laden[8].

The idea that epistemic puzzlement is theory-laden can be expressed with the concept of the *explanatory context* (hereafter, e-context)[9]. Every case of epistemic puzzlement and its corresponding why-question is couched in an e-context. A why-question's e-context consists of its theoretical and epistemic presuppositions. In the case of Jones' question regarding Roy's behavior, several such theoretical presuppositions are made. First, certain entities are invoked, such as macroscopic objects (e.g. cabs). Second, causal relationships are presupposed, namely, causal relationships between intentional states and behavior. Further, such a causal relationship itself presupposes that mental states exist.

Thus, generally, the e-context, consisting of such entities as mental states and such relations as the "causal-intentional" relation, is invoked by Jones' question, "Why is Roy hailing a cab?". In such a way, we might say that Jones' question draws from or is couched within that e-context. Further, by extension, Jones' state of epistemic puzzlement expressed by her why-question and analyzed in terms of the

---

[8] This claim is similar to assertions made by Garfinkel. He claims that questions shift or arise, depending on what one's "explanatory frame" is. The notion of the "explanatory frame" is analogous to my concept of the e-context. Garfinkel gives several examples to illustrate this point, including an example of a question that arises within an e-context of Aristotelian physics, but cannot arise on contemporary evolutionary biology (See Garfinkel, pp. 98-10).

[9] Again, the term *context* is used in a non-technical sense. The e-context is similar to van Fraassen's idea of the "background theory" of a question (147).

epistemic gap between presupposition and the set of explanatory goals of the question, is also couched within this e-context. Therefore, both the asking of a why-question and the state of epistemic puzzlement expressed in the question are theory-laden, or e-context-dependent.

Finally, it is important to note that e-contexts may vary with respect to theoretical complexity and detail. For instance, the question "Why is Roy hailing a cab?" and its subsequent answer might be drawn from folk-psychology, which is a relatively simple and imprecise e-context, consisting generally of an ontology of mental states and the causal-intentional relation. On the other hand, such a question might find a more detailed e-context, which, in addition to drawing on a folk-psychological ontology, maintains a rather sophisticated and detailed set of scientific theories. This would be the case, for example, on a Freudian-psychoanalytic explanation of a behavioral event.

Let us sum up this theory of explanation as it stands presently: Broadly speaking, explanations serve the epistemic function of removing a subject from a state of puzzlement. This puzzlement is expressed in a why-question and can be understood as the epistemic gap between the presupposition and the epistemic goal of a why-question. The subject expresses this why-question in the utterance of an interrogative sentence. An explainer, via features of the situation in which the question is asked, must ascertain the why-question expressed in the interrogative sentence. To

understand a why-question, the explainer must ascertain the topic, alternative space and relevance relation of the question.

Further, the conditions necessary for being puzzled and for asking a why-question are contained in an e-context, from which the question draws its ontology, as well as its relevance relation. Thus, the explainer, in order to understand and answer the given why-question, must, in addition to ascertaining the specific features of the why-question, grasp the e-context in which the why-question is couched. Given that the e-context supplies the features of the why-question, it also specifies the conditions by which the question will be successfully answered. Thus, in order to be relevant, an answer must be couched within the e-context in which the question was asked (i.e. make the same theoretical, ontological and epistemic presuppositions). Such is, in broad strokes, the pragmatic conception of explanation.

4. Objections and Replies

A. Objection 1: Bogus Relevance Relations

Recall that one of the problems with Kim's explanatory exclusion principle is that its application is too broad. The arguments for excluding the intentional idiom as capable of providing causal explanations can be applied to all of the

special sciences, thus making the EEP unpalatable. On the other hand, the fact that the implications of the pragmatic conception of explanation can be applied to those non-physical sciences (viz. making room for their causal laws and claims against the backdrop of physicalism) lends credence to such a theory of explanation.

However, regardless of its broad, beneficial application, there are two powerful objections typically brought to bear against the pragmatic theory of explanation. The first is given by Wesley Salmon and Philip Kitcher[10]. The core of the objection focuses on the lack of restrictions placed on the relevance relation R. According to Salmon, without such restrictions, it is not clear how to separate legitimate from illegitimate relevance relations: "If R is not a bona fide relevance relation, then [answer] A is 'relevant' to [the topic] Pk only in a Pickwickian sense." (141)

To illustrate the problem, Salmon offers an example: Smith enters an astrologer's office, places a 20 dollar bill on the astrologer's desk and utters the interrogative, "Why did JFK die on 11/22/63?" Given the clues provided by the situation, one would likely be able to extract the features of the why-question Q expressed in the interrogative: The

---

[10] This objection can be found in summary in "Four Decades of Explanation," in Minnesota Studies in the Philosophy of Science, vol. 13, 1989, and in full in "Van Fraassen on Explanation," in Journal of Philosophy, 84, pp. 315-330.

topic of Q is "JFK died 11/22/63," a likely alternative space is {JFK died 11/22/63, JFK died 1/1/63, JFK died 1/2/63, ... JFK survived 1963}, and a likely relevance relation is "astral influence."

Here, an answer must pick out the topic from the alternative space, according to the conditions specified in the relevance relation. Such an answer for Q would presumably be "a true description of the configuration of the planets, sun, moon, and stars at the time of Kennedy's birth." (142) Generally, if one were to express the interrogative, "Why did JFK die on 11/22/63?" and this interrogative expressed the question Q=<Pk, X, R>, as explicated above, it seems, in principle, possible to give a satisfactory answer to Q.

First, such an answer would be true--it would be an easy matter to chart the alignment of the planets and stars on the date of Kennedy's birth. Further, as Salmon argues, it is reasonable to assume that astrological theory will be able to show how A selects (or favors) the topic, relative to the other members of the alternative space: "Since this explanation, like any explanation, is given ex post facto, we must credit the astrologer with sufficient ingenuity to produce such a derivation." (142)

It appears possible, on the present theory of explanation, to give a satisfactory answer to a why-question whose e-context is astrological theory. But it certainly seems counterintuitive to claim that the alignment of the

planets and stars on the date of one's birth constitutes a legitimate explanation of the events in one's life. It appears that the pragmatic theory of explanation grants explanatory legitimacy to explanations which are clearly not legitimate. Again, the culprit, according to Salmon and Kitcher, is the relevance relation. If proper limitations or conditions were placed on the relevance relation R, then such explanations (involving such relations as astral influence) could not be said to bear an explanatory relation to the topic and alternative space. They would be ruled out in principle.

The burden of proof, according to Salmon and Kitcher, falls on van Fraassen to give a general account of the relevance relation which allows us to rule out illegitimate relevance relations, such as astral influence in astrological theory: "As many philosophers have insisted, we need to appeal to objective nomic relations, causal relations, or other sorts of physical mechanisms if we are to provide adequate scientific explanations." (145) In short, if we want a scientific theory of explanation, then the pragmatic theory must provide conditions by which non-scientific why-questions (viz. having non-scientific relevance relations) and their answers are excluded from consideration[11].

---

[11] Salmon and Kitcher assume that the demarcation between legitimate and bogus explanation is science. A good theory of explanation will, ultimately, be a good theory of scientific explanation. While it is not clear to me that this is true, I will not dispute the point, since the central concern here is to legitimize explanatory claims made by a science of the mind.

However, as Salmon and Kitcher point out, determining what exactly those conditions are is a matter of extreme difficulty. Let us explicate the problem using the vocabulary we have developed thus far. On the present theory, why-questions draw relevance relations from the e-contexts in which they are couched. Again, an e-context can be understood, in most general terms, by its ontology and relevance relations. As such, astrology forms an e-context for the above why-question, expressed in the interrogative, "Why did JFK die on 11/22/63?" But it is highly doubtful that e-contexts such as astrology provide bona fide relevance relations.

The problem is, on the pragmatic theory of explanation, there are no mechanisms by which such bogus relevance relations (or, more generally, bogus e-contexts) are excluded as incapable of providing legitimate explanations. The topic is true and, within the e-context of astrology, an answer A which bears the relevance relation "astral influence" to the topic and alternative space is forthcoming. After all, say Salmon and Kitcher, "to the sincere believer in astrology the configuration of heavenly bodies at the time of Kennedy's birth is highly relevant to his death on that particular fateful day in 1963." (145) E-contexts which invoke astrological theory appear to relieve the puzzlement of those subjects who believe in astrology. Thus, the pragmatic theory of explanation appears to include explanations that are illegitimate. There are no devices built into the theory

which exclude obviously bogus explanations and explanatory contexts, nor are there any such devices readily forthcoming.

On such a theory of explanation, how does one differentiate between scientific and non-scientific explanations? Van Fraassen anticipates this type of objection and replies:

> To call an explanation scientific, is to say nothing about its form or the sort of information adduced, but only that the explanation draws on science to get this information (at least to some extent) and, more importantly, that the criteria of evaluation of how good an explanation it is, are being applied using a scientific theory. (155-156)

To put this response in the present terminology, whether or not an explanation (as an answer to a why-question) is legitimate (i.e. scientific) is not determined by any conditions placed on any feature of the explanation, including the relevance relation. On the contrary, whether or not an explanation is scientific is determined by whether or not the e-context, from which both the question and the answer are drawn, is regarded as a scientific e-context. Specifically, is the theory (or theories) which couches the why-question and its answer-candidates, and thus forms the question's e-context, regarded as a scientific theory?

Salmon and Kitcher acknowledge this response, but seem to find it inadequate, although their reasons are somewhat obscure. It appears that their rejection of this reply is based in the idea that an adequate theory of explanation will be a theory of scientific explanation. They require that the

theory of explanation itself set up conditions by which we can judge explanations scientific or non-scientific:

> Unless we can impose the demand for objective relevance relations, we cannot arrive at a satisfactory characterization of scientific explanation ... We need to appeal to objective nomic relations, causal relations, or other sorts of physical mechanisms if we are to provide adequate scientific explanations. (145)
>
> We must be quite clear that it is scientific explanation with which we are concerned. The term "explanation" is used in many ways that have little or nothing to do with scientific explanations. (6)

Van Fraassen's response to Salmon and Kitcher's objection is that it is not the job of a theory of explanation to judge whether or not a particular explanation is scientific. Such a judgment is made of the e-context itself: Is the e-context, in which the why-question and its answer are couched, a scientific e-context? Is the e-context constituted of presently-accepted scientific theories? Salmon and Kitcher seem to reject this response, since (as they see it) it does not provide them with an adequate theory of scientific explanation. Such a theory should be able, on its own, to exclude non-scientific explanations.

But such a requirement must be in vain. Such a condition on the relevance relation, which excludes, for instance, astral influence as a bona fide relevance relation, must be misguided, since astrology might very well have been true. It is well within the realm of possibility that there be a causal relation between the "alignment" of the stars and

planets at the time of one's birth and the resulting course of one's life[12].

Further, such a requirement presupposes that what counts as good science is static. If the conditions by which an explanation is to be judged scientific must be contained in the theory of explanation itself, instead of simply requiring explanations to draw from currently accepted scientific theories, then what counts as good scientific theory cannot change, without changing the theory of explanation.

But certainly what counts as good science does change, without a corresponding change in our theory of explanation. Such theorists as Thomas Kuhn and Larry Laudan have pointed out that what counts as good science is, at a variety of levels, subject to change:

> Science offers the remarkable spectacle of a discipline in which older views on many central issues are rapidly and frequently displaced by newer ones. ... Moreover, change occurs at a variety of levels. Some of the central problems of the discipline change; the basic explanatory hypotheses shift; and even the rules of investigation slowly evolve" (Laudan, 4-5)

Kuhn's idea of the scientific revolution as a shift between incompatible scientific paradigms implies the possibility of change in what counts as good science. Theories such as Laudan's and Kuhn's illustrate the dynamic

---

[12] Perhaps, such a causal relation would qualify as action at a distance, but the possibility of such causal relations has arisen in contemporary quantum mechanics. Action at a distance, therefore, does not exclude such causal relations as impossible.

nature of scientific investigation. What counts as good scientific theory, and thus good scientific explanation, is subject to change. Hence, any attempt to place restrictions on relevance relations will be in vain, since a change in what counts as good science might well involve a violation of such restrictions.

On the pragmatic theory of explanation, there are two forums for judging the adequacy of explanations. First, an answer-candidate is judged to be adequate at the level of the e-context in which the why-question is couched. Questions of explanatory adequacy that take place within the e-context, according to the conditions established by the why-question via the relevance relation, can be called (in Carnap's terms) "*internal*" questions[13].

Regardless of whether the answer-candidate passes muster from within the e-context, one might question the legitimacy of an explanation by challenging the relevance relation of the why-question, to which the explanation is an answer. A challenge to the e-context of astrology as incapable of providing adequate explanations would fall into this category. Such questions which challenge the legitimacy of

---

[13] While the "internal/external" distinction used here closely resembles Carnap's distinction, they are not identical. Cf. "Empiricism, Semantics, and Ontology" in Meaning and Necessity (1956: University of Chicago Press) pp. 205-221.

the e-context in which explanations are given can be termed *"external"* challenges or issues[14].

This is, I believe, the basis of van Fraassen's reply to this type of objection and I believe it is fundamentally correct. The pragmatic theory of explanation is a theory of explanation in general. It does not have the resources to adjudicate between scientific and non-scientific explanations, nor ought we expect it to. That is not its job. To question whether an explanation is scientific is to challenge the theories of the e-context in which why-questions and their subsequent answers are couched. Salmon and Kitcher are correct: We need to be able to adjudicate between scientific and non-scientific explanations. But such adjudication does not (and cannot) take place within the theory of explanation.

B. Objection 2: Explanatory Relativism

A second common objection to pragmatic theories of explanation regards the central role played by the subject in such theories. The importance of the subject in explanation has long been recognized, but most theorists attempt to downplay or eliminate the role of the subject from theories

---

[14] We might ask such an external question of the intentional idiom. Is the intentional language a legitimate (i.e. scientific) e-context or is it more closely akin to the likes of astrological theory? This question shall be addressed in chapter IV.

of explanation. The problem, as philosophers such as Hempel and Gasper show, is relativism:

> To explain something to a person is to make it plain and intelligible to him, to make him understand it. Thus construed, the word 'explanation' and its cognates are pragmatic terms: their use requires reference to the persons involved in the process of explaining. In a pragmatic context we might say, for example, that a given account A explains fact X to person P_1. We will then have to bear in mind that the same account may well not constitute an explanation of X for another person P_2. ... Explanation in this pragmatic sense is thus a relative notion: something can be significantly said to constitute an explanation in this sense only for this or that individual. (Hempel, 425-426)

Philip Gasper shares Hempel's concern over making sense of the role of the subject in explanation:

> A satisfactory explanation of an event or phenomenon should provide us with understanding of what has been explained (the "explanandum"). But understanding is a notoriously vague and subjective state. Different inquirers may disagree about what is sufficient for understanding and whether or not understanding has actually been achieved. If the search for explanation has a central role in scientific reasoning, then it is important to ensure that our concept of explanation is free from this kind of vagueness. (Gasper, 289)

Does taking the subject to be a central feature in explanation theory entail that explanatory adequacy will be relative to the subject seeking the explanation? If so, then the pragmatic theory may be lost in radical relativism, which will effectively preclude its being able to provide an analysis of objective explanation. The question is, what exactly is the role of the subject in determining explanatory adequacy and does such a role entail relativism?

On the present analysis of the pragmatic theory of explanation, there are two ways in which the subject seeking the explanation plays a crucial role. First, *an explanation is adequate only if it addresses the reason or reasons for which the subject seeks an explanation.*

Recall, this was the difficulty illustrated by the example in which both Smith and Jones inquire after Roy's arm-movement behavior. An answer to Jones' question, "Why is Roy hailing a cab?", based in the physical idiom is inadequate since it doesn't address the reason or reasons for Jones' asking the question. It will not move the subject from the presupposition to the epistemic goal of the question. The conditions by which such movement is accomplished are determined by the features of Jones' why-question, including the question's relevance relation--the "causal intentional" relation. A physicalist answer fails to remove the subject, Jones, from her state of puzzlement, so it is irrelevant.

Notice, however, that while the subject plays a crucial role in explanatory adequacy, in that the answer must be relevant to the subject's question, the criteria by which an answer can be said to be relevant (i.e. address the reasons for which the subject seeks an explanation) is determined by the relevance relation. The relevance relation is drawn from a particular e-context. Again, the legitimacy of the relevance relation can be scrutinized by asking whether the e-context from which it is drawn consists of bona fide

scientific theory (this is the "external" question). E-contexts and their relevance relations are not subjective. They are theoretical entities subject to public criticism. Since it is the e-context that couches the why-question and it is the features of the why-question which determine the conditions by which a satisfactory answer can be given, such conditions are subjective only if e-contexts are subjective. But they are not.

Explanations serve the epistemic goals of the subject. Thus, the subject selects the why-question and so determines the e-context from which the question is both understood (both by explainer and explainee) and answered. However, once the selection of the e-context is made, the conditions by which an answer is judged adequate are determined by the e-context in which the why-question is couched. The subject selects an e-context, but it is the e-context that determines the conditions for explanatory adequacy (viz. the relevance relation). These conditions, in turn, are subject to external public scrutiny. Thus, there is no relativism in the conditions of explanatory adequacy.

Explanatory adequacy is objective in the sense that scientific e-contexts are public domains of inquiry. The theories, entities and relevance relations that make up an e-context are not private languages. Thus, the conditions for explanatory adequacy, determined by the e-context, cannot

vary from subject to subject[15]. Which e-context is selected may vary from the subject to subject, but once the selection is made, the conditions for explanatory adequacy are determined by the elements of the e-context. Such a role for the subject does not entail relativism for the pragmatic theory of explanation.

The second manner in which the subject plays a crucial role in explanation theory tends to be a bit more tricky: *An answer must not go beyond the scope of the subject's knowledge*. An answer might be relevant in the sense that it meets the conditions specified in the relevance relation, but still go beyond the scope of the subject's knowledge. Thus, an answer might meet all of the conditions for adequacy, but still not move the subject to the explanatory goals of the question.

For instance, a first-year student of physics asks a professor to explain why objects cannot move faster than the speed of light. Clearly, the e-context from which the question is asked involves theories of space-time physics and the professor is able to answer this question. However, the professor knows that the best possible answer, with all the mathematics involved, will not be grasped by a first-year student of physics. The student has enough of a grasp on the

---

[15] This is not to say that there is no room for dissent in publicly constructed theories, such as those theories found in science. Lauden gives a very nice account of dissent in the scientific community, as well as the process by which consensus is formed. See Lauden, Science and Values, U. California Press: 1984.

e-context to be able to ask the question, but not enough to understand the best possible answer.

The problem is, simply because a subject is able to ask a question from within a particular e-context does not imply that he has enough of a mastery of that e-context to understand an appropriate answer. Certainly, there are varying degrees of understanding one might have, with regard to an e-context. Thus, it appears that the conditions for understanding an answer will vary from subject to subject, depending on the level of mastery a subject has over the e-context in which the question is couched.

This problem illustrates that maintaining the subject as a central feature of explanation theory makes such a theory complicated, but it does not introduce an "anything goes" brand of relativism into the picture. In most cases, an explainer is able not only to ascertain the e-context from which the question is asked, but also the level of mastery the subject has over that e-context. The physics professor, for instance, recognizes that the subject desiring the explanation is a first-year student and thus does not as yet have the background to understand the most complete, sufficient answer to the question.

However, an explainer can give an answer which, although not the best answer possible, still selects (or favors) the topic from the alternative space in the manner specified by the relevance relation. The answer draws on the ontological and theoretical machinery of the e-context, but in a broader,

more vague manner. This type of explanation occurs frequently, and not simply in academic settings. Doctors, for instance, answering questions such as, "Why do I need surgery?", are able to give an adequate answer without invoking the technical vocabulary necessary to give the best answer.

This may open the door to a variety of possible answers, each varying from subject to subject with respect to complexity and detail, but we are able to avoid relativism in two ways: First, all the possible answers still draw on the e-context from which the question is asked and, thus, they meet the conditions by which explanations are judged adequate, which are determined by the e-context, rather than the subject. Second, in cases where answer A is regarded as the best answer, but the subject does not understand answer A, we can still claim that, *where A is the best answer to question Q, within e-context C, if subject S were to have a sufficient mastery of e-context C, from which question Q is asked, then he would understand answer A.*[16] In cases where a subject's puzzlement is not removed because the subject lacks the requisite mastery of the e-context, and thus does not understand an answer, it can still be claimed that he would understand the answer if he had sufficient mastery of the

[16] Notice, this is not intended to be a definition of the best answer. This claim merely allows us to conceive of the idea of A being the best answer to question Q (relative to the e-context), even though subject S does not understand A. Such an answer would be appropriate if S were to have sufficient mastery of the e-context of Q.

e-context which couches the answer.

Therefore, we are able to maintain this second subjective condition of explanation, namely, that an answer must not go beyond the scope of the subject's knowledge, without committing ourselves to a relativism of the conditions by which answers are judged adequate (or better or worse). So the subject can maintain a central role in the pragmatic theory of explanation without such a theory finding itself mired in relativism.

5. Example and Summary

Before we move on, let us sum up the theory as it stands. The epistemic function of explanation is to remove a subject (i.e. an explainee) from a state of epistemic puzzlement. This state of puzzlement is expressed in a why-question, in which puzzlement is understood as the epistemic gap between the presupposition and the epistemic goal of the question. An explainer, in order to remove the subject from his state of puzzlement, must understand the why-question expressed in an interrogative sentence. Such an understanding is accomplished, via clues provided by the situation in which the question is asked, when the explainer ascertains the topic, alternative space and relevance relation of the subject's why-question.

However, epistemic puzzlement cannot occur in a theoretical vacuum. Ontological, theoretical, and epistemic presuppositions are necessary for the possibility of epistemic puzzlement. These presuppositions, expressed in the above features of why-questions, draw from a particular set of theoretical commitments--commitments which can range from extremely broad, such as general relevance relations and ontology, to extremely detailed, such as specific scientific theories.

Such theoretical commitments form the e-context of the why-question. The e-context, as well as the specific features of the subject's why-question, can be ascertained by clues provided by the situation in which the question is asked. Answers to why-questions are explanations and such answers, in order to succeed as explanations, must be relevant to the question asked. The conditions for explanatory relevance are determined by the relevance relation of the question. Such relevance relations are drawn from the e-context in which the question is couched. Thus, in order to be relevant, an answer must ultimately be relevant to the e-context in which the why-question is couched. Answers that are not relevant to the e-context of a question fail, since they do not relieve the subject's puzzlement (i.e. they do not move the subject from the presupposition to the epistemic goal of the why-question).

At this point, it would be instructive to demonstrate how an example from scientific practice fits this model of

explanation. Such an example will be helpful for two reasons: First, by being grounded to a particular situation, it can be shown that the pragmatic theory of explanation is able to demonstrate its applicability--that it is useful in understanding how explanations work in scientific practice[17]; second, an example will show that seeking and finding explanations is often a rather complicated procedure. Such complications, as will be shown, can be handled on the pragmatic model.

The pragmatic theory presents a sort of story about when explanations are required and when those requirements are met. The story begins with a subject encountering a phenomenon. That phenomenon is a source of puzzlement for the subject, in that she knows *that* the phenomenon occurred, but not *why*. The search for an explanation is a search for a resolution to this puzzlement. The end of the story is generally the resolution of the puzzlement through the acquisition of an explanation of the phenomenon. Such a story appears linear. We begin with puzzlement and end with resolution.

---

[17] That is not to say that the pragmatic theory is universally applicable, nor is its value found purely in its application. It is not yet clear that pragmatic theory can account for every explanation we regard as successful. However, even if there are anomalies that pose challenges to the present theory (and I'm sure there are), its value is at least shown in allowing philosophers to cope with situations where there are multiple complete and independent explanations of a single explanandum event. As Kim argues, it is not clear that the standard D-N models of explanation can cope with such a phenomenon.

In science, however, the situation can be more dynamic. Often, scientists have answers (in the form of hypotheses about the world) but search for questions to ask. In other words, a scientist's theory or hypothesis is vindicated by how well it solves problems or resolves puzzlement. The more questions a theory or hypothesis can answer, the more valuable it is. Therefore, it often occurs in science that the story is backwards--the process moves from answer to question. Richard Feynman phrases it this way: "We physicists are always checking to see if there is something the matter with the theory. That's the game, because if there is something the matter, it's interesting!" (Feynman, 8) Scientists seek to be puzzled. Once they are puzzled, they can discover how well their theories work, or how much work their theories need. Let us take an example, of such a situation and show how it fits the pragmatic model of explanation[18].

In the early 19th century, physicists hypothesized (had a hunch, really) that the phenomena of electricity and magnetism were somehow related. Both phenomena were well documented and understood, but there were as yet no demonstrations of any relation between the two. As George Gamow says, "Electric charges did not influence ... magnets in any way; neither did ... magnets influence ... electric

---

[18] The following example comes from George Gamow's The Great Physicists from Galileo to Einstein, (Dover Publications: 1961) pp. 135-137.

charges." (Gamow, 135)

It is the physicist Hans Christian Oersted who is credited with discovering the phenomenon that would lead to the development of electromagnetic theory--the unified theory of magnetism and electricity. Oersted's experiments involved an electric current generated by a Volta pile[19] and a compass. Similar experiments prior to Oersted's generally involved interaction between a compass and a static electric charge. Instead of using static electricity, Oersted connected a wire to the two poles of the Volta pile, thus creating a directional current of electricity. He then placed the compass near the wire: "The needle, which was supposed to orient itself always in the north-south direction, turned around and came to rest in the direction perpendicular to the wire." (ibid. 135) When the Volta pile was disconnected, the compass again pointed towards magnetic north. Once the wire was connected to the Volta pile, the compass again turned away from magnetic north to become perpendicular with the wire.

Here was an anomalous phenomenon. Present, accepted theories of electricity and magnetism could not account for it; for, at this point, electromagnetic theory was only a hunch, much less a fully developed scientific theory. Hence, Oersted found himself in a state of puzzlement, expressed by

---

[19] A "Volta pile" is basically a battery, consisting of "alternating copper and iron or zinc discs, separated by layers of cloth soaked in salt solution." (Gamow, 133)

the question, "Why did the compass needle turn perpendicular to the wire?". The topic is the event of the needle turning perpendicular to the wire, away from magnetic north. Given what Oersted knew about magnets, a reasonable alternative space for the question would consist of {the needle turns perpendicular to the wire, the needle remains pointing towards magnetic north}. Thus, a successful answer must select the topic from the alternative space, demonstrating why the needle turns perpendicular to the wire, as opposed to pointing at magnetic north, as is expected.

At this point, Oersted's hunches about the relation between electricity and magnetism are rather vague (since his was the first proof that there *was* such an interaction). So the relevance relation would be a broad causal relation: something is causing the compass to turn away from magnetic north.

Therefore, for question Q, "Why did the compass needle move perpendicular to the wire?", Q is identical to the ordered triple <X, Pk, R>, where the topic, X, is the event of the needle turning perpendicular to the wire attached to the Volta meter, the alternative space Pk is {the needle moves perpendicular to the wire, the needle remains pointing at magnetic north} and the relevance relation R is a causal relation having to do with the electricity.

Recall that physicists at this time had a broad, vague hunch that there was some sort of relationship between magnetism and electricity. In order to resolve Oersted's

puzzlement regarding the phenomenon of the compass, however, that hunch would be honed into a broad hypothesis:

> It became quite clear to [Oersted] that there was an interaction between the magnets and the *moving* electricity, and the direction in which the compass needle was oriented depended on the direction in which the electrical current was flowing through the wire. (ibid. 136)

The e-context of the question, "Why did the compass needle move perpendicular to the wire?", consists in this case of the epistemic presupposition of the compass' movement, the theories of electricity and magnetism, as well as certain hunches about the interaction of their respective phenomena.

Notice that this case demonstrates the claim that puzzlement cannot occur in a vacuum: The event of the compass' moving to become perpendicular with the wire is puzzling only if it somehow confounds our expectations about how the magnet *should* have behaved. Knowing what we do about magnetic theory, the unfettered magnet was *supposed* to point toward magnetic north. That expectation, founded in our knowledge of how magnets work, is foiled. That is what makes the event puzzling and, according to Feynman, exciting.

Thus, puzzlement, expressed in our why-question, is couched within the e-context containing electrical theory (which allows us at least to understand terms like "current") and magnetic theory (which allows us at least to form foiled expectations about the behavior of the magnet). This

e-context also contains the epistemic presupposition of the activity of the magnet and the relevance relation of the causal interaction between electricity and magnets, which is itself couched in the hunch of the physicists of the day.

This example demonstrates not only how puzzlement, expressed in the form of a question, must be couched in an e-context, but that the relationship between questions and answers is dynamic. First, scientists do not necessarily begin with puzzlement and seek its resolution in accepted theory. Both new and old theory find vindication in the resolution of puzzlement. Thus, scientists actively seek to be puzzled, not only to see if their present theories cut the mustard, but also to develop new theories or to test and hone hypotheses or hunches.

Further, this example demonstrates a dynamic relationship between the arising and alleviation of puzzlement. Once Oersted's original puzzlement is resolved, new questions and puzzlements arise. For instance, Oersted might ask, why did the compass turn perpendicular to the electrical current, as opposed to spinning like a top, or turning parallel to the electrical current, and so on. While old puzzlements are resolved, new puzzlements arise. In this dynamic interaction between why-questions and answers within an e-context, theories are developed, honed, and vindicated. This example demonstrates that the pragmatic theory of explanation can not only make sense of concrete cases of explanation-seeking, but it can also tell a story about the

rather dynamic relationship between why-questions and answers in scientific inquiry.

The task at this point is to show, in more detail how such a theory of explanation allows us to understand the role of the mental in the production of behavior. This is the aim of chapter IV. Specifically, we shall focus on what type of analysis of mental causation follows from the pragmatic theory of explanation. Further, we shall entertain an additional objection to the idea that the mental plays a systematically irreducible role in the production of bodily behavior.

Chapter IV: Mental Causation and Explanation

1. Problems and Resolutions

In order to illustrate how an analysis of mental causation can be generated from the pragmatic model of explanation, let us briefly return to the original formulation of the problem. Against the backdrop of physicalism, there is a fundamental tension between the irreducibility and the causal efficacy of the mind. The principles of physicalism dictate that mental states and their properties are determined by physical states of the body. However, given arguments that demonstrate the irreducibility of mental states (such as multiple realizability) such mental states, properties and powers cannot be systematically identified with their physical realizations. These claims are expressed in the DBNI thesis, where the properties of mental states are "determined by but not identical to" physical states.

The causal power of mental states is an example of a type of mental property that can be subsumed under the DBNI thesis. To claim that a mental state M is sufficient for the production of bodily behavior E implies that the physical realization P of M is also a sufficient cause of E; for, if M were sufficient for E and P were not, then M would have a causal property that P does not, namely, being sufficient for E. This would contradict the DBNI thesis. Thus, one

encounters a situation in which there are two distinct events, M and P, each of which is a sufficient cause of E.

On such a view, at least two problems emerge. First, if P is a sufficient cause of E, then it is not clear what causal work remains for M to accomplish. How does one make sense of M as a cause of E? What work is there for M to do that is not already accomplished by P and how is such work to be distinguished from that done by P?

A second, stronger problem is causal overdetermination. Recall that causal overdetermination does occur, but cases of overdetermination present special problems for causal analysis. If M and P are each separate, sufficient causes of E, then it is unclear how a causal analysis of E should proceed. Since both M and P are sufficient causes, each renders the other unnecessary. We cannot, therefore, claim that P caused E, since E would have occurred whether P had happened or not. P is superfluous to E. The same claim can be made of M. If M and P are each superfluous to E, then it is unclear what caused E. E has a cause, but a causal analysis of E is at best problematic.

Again, instances of causal overdetermination do occur, but as oddities or anomalies. But according to the DBNI thesis, *every* intentional act is causally overdetermined. In other words, the DBNI thesis implies systematic overdetermination, thus rendering causal analyses of *all* intentional behavior problematic. A primary goal of contemporary philosophy of mind, recall, is to assimilate the

mind into our scientific world-view. But one of the central goals of science is to ascertain and make sense of the causal structure of the world. To conceive of the mind as systematically irreducible to the brain, and yet causally responsible for behavioral events appears counterproductive to that goal, in that an understanding of the causes of "intentional" behavior is made obscure and problematic by systematic overdetermination. Therefore, if the DBNI thesis does indeed imply systematic overdetermination, then it is at best problematic and at worst absurd.

So any cogent analysis of mental causation must meet two goals: First, such an analysis must make room for and make sense of the unique causal work done by the mental in the production of behavior, while retaining the claim that physical events are capable of providing sufficient causes of such behavior; second, it must avoid the pitfall of systematic overdetermination. The pragmatic theory of explanation provides an analysis of mental causation which achieves both of these goals.

First, how is the unique causal work of the mental to be understood, given that the physical is capable of providing sufficient causes of any behavioral event? An analysis of and answer to this first problem (which was begun in chapter III) can be given by the pragmatic theory of explanation. Recall that the implications of physicalism, namely, that any behavioral event is subject to an exhaustive causal analysis within the physical idiom, can be understood to mean, first,

that for any behavioral event, there is a why-question regarding that event which requires an answer from within the physical idiom; second, any such question can be given a correct answer from within that idiom.

What is not implied by physicalism is that, for any behavioral event, *any* why-question regarding that event can be answered in the physical idiom. Why-questions whose relevance relations are causal-intentional, for example, demand an answer/explanation of a behavioral event which refers to the mental states of the subject performing the behavior. Such is the case with the example of Roy's hand-waving. Jones' question, "Why is Roy hailing a cab?" characterizes Roy's behavior as purposive or intentional, thus requiring an explanation which refers to his mental states.

The unique causal work done by mental states in the production of behavior can be understood through an analysis of the role such states play in explanations of behavior. The causal analysis of behavioral events which involve irreducible mental states is capable of providing answers to why-questions which cannot be answered by reference to physical states. In other words, in particular sorts of cases, the mental idiom is capable of providing causal explanations where the physical idiom fails to do so.

One might object that this seems to imply a sort of explanatory failure for the physical idiom. One of the implications of physicalism is that the physical idiom is not

subject to explanatory failure. That is, the principles of physicalism dictate that every behavioral event has a sufficient physical cause. So, hasn't the pragmatic theory of explanation, as an analysis of the role of the mental in causing behavior, somehow violated the principles of physicalism?

On the contrary, this analysis of mental causation suggests that there are two distinct types of explanatory failure. By way of demonstration, let us return to the example of Roy's hand-waving. Recall Jones' question, "Why is Roy hailing a cab?", where such a question has a causal-intentional relevance relation. On the other hand, Smith's question is "Why is Roy's arm moving about?", where such a question has a causal-mechanical relevance relation.

Suppose, first, that Roy's behavior is not intentional--that it is a spasm or seizure caused by a neurological condition. In this case, any intentional characterization of Roy's behavior (as the topic of a question) will be false. Recall, however, that in order for a question to be answerable (i.e. to admit of a right answer) its topic must be true[1]. So, in the present scenario, any why-question who's topic or presupposition characterizes Roy's behavior as intentional cannot be answered since such a characterization is false.

---

[1] Similarly, in Hintikka's terminology, the claim can be made that in order for a wh-question to admit of a correct answer, it must have a true presupposition. See chapter II, p. 8.

In this example, there is a failure to explain in a rather broad sense. The failure is a sort of breakdown of the e-context itself. The intentional e-context (maintaining an ontology of mental states and the causal-intentional relevance relation) is incapable of describing Roy's behavior as the topic of a why-question. Thus, any such why-question couched within the intentional e-context cannot be answered, since its topic is false. The e-context itself is incapable of formulating an answerable why-question regarding Roy's behavior, since it is incapable of describing his behavior. Therefore, no explanation of Roy's behavior can be given from this e-context. Such broad explanatory failure we might call a ***failure of description***, since the e-context is incapable of describing the event.

On the other hand, suppose that Roy's behavior *is* subject to an intentional description. In particular, let us say that the topic, "Roy is hailing a cab" is true. Since the topic (and hence the presupposition of the why-question, "Why is Roy hailing a cab?") is true, the intentional e-context is (ceteris paribus) capable of providing an explanation of Roy's behavior. Since the relevance relation of the question is causal-intentional, any non-intentional answer, including any answer from the physical idiom, will be irrelevant. Here, the physical idiom fails to explain in that any physicalist answer to the question "Why is Roy hailing a cab?" will be irrelevant. This narrower type of explanatory failure can be labeled, ***failure of relevance***.

Physicalism implies that the physical idiom cannot be subject to a failure of description. For any behavioral event, a why-question, couched in the physicalist e-context, whose topic is true, can be asked and answered. Such a claim cannot be made for the intentional idiom. Thus, the intentional idiom can be subject to a failure of description. The physical idiom, however, is not subject to such failure.

While physicalism implies that the physical idiom is not subject to a failure of description, there are no such implications of physicalism for failure of relevance. The claim that there are cases in which any answer from the physicalist e-context will be irrelevant to a question asked from the intentional e-context is compatible with physicalism, since a failure of relevance does not imply a failure of description.

Hence, there are two general types of explanatory failure that involve a failure of the e-context itself[2]. First, if e-context A is incapable of providing an answer to question Q because it cannot provide an answer which meets the conditions of the relevance relation of Q, then A has suffered a ***failure of relevance***. Since A is incapable of providing an answer to Q, it is the e-context itself that has failed.

---

[2] There is, of course, the possibility that an explanation could fail because it is false. This type of failure, however, does not entail a failure of the e-context. It simply implies that the answer given is not correct.

Secondly, if e-context A is incapable of describing or characterizing an event as the topic of a question, then the e-context fails to explain in that it cannot formulate an answerable question which has that event as its topic. This too is a failure of the e-context. Such a failure is a ***failure of description***, since the e-context is incapable of describing the event as the topic of a why-question.

Thus, we are able to answer the first problem of mental causation. First, as was argued in chapter II, causation can be understood and analyzed in terms of explanation. 'A causes B' can be understood in terms of 'A explains B.' Second, the pragmatic theory of explanation allows us to understand the unique causal work done by the mental. In other words, the pragmatic theory allows us to make sense of the causal efficacy of the mental without violating the principle of physicalism. For on the pragmatic theory, a causal analysis of a behavioral event which involves mental states can provide causal explanations that a physical analysis cannot. A why-question whose relevance relation is causal-intentional demands an answer which draws from the mental states of the subject performing the behavior. Since the physical idiom is in principle incapable of answering such a question (i.e. the physical idiom is subject to a failure of relevance), the mental idiom is capable of providing causal explanations of behavior that the physical idiom cannot. In this way, we are able to understand the

unique work done by the mental in production of behavior. Further, since such a failure of relevance by the physical idiom does not imply a failure of description, this analysis of the role of the mental in the production of behavior in no way violates the principles of physicalism.

But what of the second problem; namely, that of causal overdetermination? Let us suppose a behavioral event E, which is subject to an intentional explanation M and a physical explanation P, both of which are seen as irreducible and sufficient in the production of E. The problem in cases of overdetermination is that there is causal competition between M and P. Since each is a sufficient cause of E, each renders the other unnecessary for E, making a causal analysis of E highly problematic.

But how is causal overdetermination, and thus this notion of causal competition, to be understood within the pragmatic theory of explanation? Recall Kim's notion of causal overdetermination: Overdetermination occurs when there are two causal explanations, P and M, for the same event, E[3]. On the present analysis of explanation, E is an event to be explained in virtue of its being the topic of a particular why-question. The parameters of the why-question (i.e. the alternative space and the relevance relation) determine the conditions by which an explanation is deemed successful.

---

[3] The fact that M and P are each complete and independent explanations of E is presupposed. In other words, neither M nor P reduce to the other.

Thus, to claim that M and P are each successful explanations of E is to claim that they each satisfy the conditions specified by the why-question. In other words, in order for M and P to be successful explanations of E, they would each have to be successful answers to a particular why-question, whose topic is E. Therefore, on the pragmatic theory of explanation, to have multiple explanations of a particular event is to have multiple satisfactory answers to a particular why-question.

In such a way, we are able to understand the causal competition inherent to overdetermination: Multiple explanations compete only if they are each successful answers to the same why-question. Again, overdetermination involves some sort of causal competition. On the pragmatic theory, there is such competition only if there are multiple answers to a particular why-question from within the same e-context. Such is the case with Kim's examples:

> A man is shot dead by two assassins whose bullets hit him at the same time; or a building catches fire because of a short circuit in the faulty wiring and a bolt of lightning that hits the building at the same instant. It isn't obvious in cases like these just how we should formulate an explanation of why or how the overdetermined event came about. (Kim II, 252)

In each example, there are multiple causal analyses of a particular event. But on the pragmatic theory of explanation, having multiple causal analyses does not, by itself, entail overdetermination. There is causal competition amongst these analyses only if they each serve as

answers to a single why-question which has the event in question as its topic.

"Why is this man dead?" can be sufficiently answered by reference to either assassin's bullet. "Why did the house burn down?" can be sufficiently answered by reference either to the lightning strike or to the faulty wiring. In each example, there is causal competition because there are multiple answers to each question. Each causal analysis, since each is mutually exclusive, rules out the other.

Broadly speaking, such causal competition occurs when an e-context offers multiple answers to a particular question. Thus, on the pragmatic theory of explanation, causal overdetermination can be understood as having multiple answers, within a particular e-context, to a particular why-question.

Further, this analysis of overdetermination can also make sense of Kim's claim that causal overdetermination produces an "epistemic predicament" for the subject seeking an explanation of an event. Recall Kim's argument: A subject is, first, in such a predicament with regards to a particular event. An explanation of the event will relieve the subject of his or her predicament. However, where such an event is causally overdetermined, not only does the original predicament remain, but a new type of predicament emerges: Which of the two explanations of the event is true? So, according to Kim, there is an epistemic problem with overdetermination. Instead of relieving them,

overdetermination compounds a subject's epistemic predicaments about the world, which is counterproductive to the epistemic goals of explanation.

On the pragmatic theory of explanation, the subject expresses her epistemic puzzlement through a why-question. In situations of causal overdetermination, there are multiple, sufficient answers to that why-question. Kim is, therefore, correct: Having multiple answers to a particular why-question *is* counterproductive to the goals of explanation: Not only does the original why-question remain unanswered, but a new question arises, namely, "Which answer is the correct one?" In this way, puzzlement is compounded.

However, such puzzlement is compounded only if there are multiple answers to a particular why-question. The present analysis of mental causation avoids this problem. The mental and the physical answer different why-questions about behavior. Thus, puzzlement is not compounded. So, although Kim is correct about the epistemic problems of causal overdetermination, such problems do not arise on the present analysis of mental causation.

If this analysis of mental causation is accurate, then there is no causal competition between mental states and physical states in the production of behavior. Each causal analysis does different causal work in that each is capable of answering different why-questions. Where a mental explanation answers such a why-question, a physical explanation is irrelevant. Where a physical explanation

answers such a why-question, a mental explanation is irrelevant. Each answer is given in a different e-context and neither answer makes any commitments about the other.

Hence, the pragmatic theory of explanation is able to make sense of the anomalous cases of causal overdetermination, as well as the difficulties such cases present for explanation. Further, the same theory shows that mental causation does not imply systematic overdetermination. By generating an analysis of the unique work done by the mental in the production of behavior and by showing how mental causation avoids systematic overdetermination, the pragmatic theory of explanation is able to make room for and make sense of the causal efficacy of systematically irreducible mental states.

2. Justification of the Relevance Relation

It is the unique character of the mental that allows the pragmatic theory of explanation to make sense of the causal work done by the mental and to show how the problem of systematic overdetermination can be avoided. But in the solution to these problems, there has been a major presupposition; namely, that the mental can be regarded as causally efficacious at all.

The primary goal of this thesis is to make sense of and make room for the causal efficacy of irreducible mental states, given that the physical sciences can, in principle, offer an exhaustive causal analysis of any behavioral event.

To this end, it has been presupposed that there *is* such a bona fide causal relation between mental states and physical behavior. However, the argument can be advanced that such a presupposition is, itself, unwarranted.

Recall the problem posed by Salmon and Kitcher for the pragmatic theory of explanation: Without a priori conditions on the relevance relation of a why-question, how are we able to discern genuine inquiry, such as that found in the e-contexts of science, from the illegitimate type, such as that found in astrology? In other words, by what conditions can genuine relevance relations, such as the causal-mechanical relation, be demarcated from bogus relevance relations, such as astral influence? The reply given in chapter III is that it is not the task of a theory of explanation to discern genuine from bogus relevance relation. On the contrary, such a problem is an external concern, which investigates the legitimacy of the e-context itself--whether or not the e-context is capable of providing adequate relevance relations.

But such a response implies another objection: If inquiries into the legitimacy of relevance relations are external questions, then such a question requires asking: Is the intentional idiom capable of providing genuine relevance relations? In particular, is the causal-intentional relevance relation a bona fide causal relation, or is it to be relegated to the likes of astral influence. Is the mental genuinely efficacious, or is mental causation an anachronism with which science will eventually dispense?

While it is not the central concern of this thesis to ask and answer this external question, it appears necessary to offer strategies by which one might demonstrate the legitimacy of the causal-intentional relation. Philosophers such as Jerry Fodor have offered arguments in support of the view that the mind bears a causal relation to certain cases of bodily behavior. In order to dispel qualms about the legitimacy of the causal-intentional relevance relation, these arguments will be discussed briefly.

3. Fodor and Ceteris Paribus Laws

As noted in chapter I, the question of causal efficacy is not simply about the relation between events, i.e. whether one event is causally responsible for another. The question is about properties of events. A red ball, for instance, collides with a blue ball with a certain amount of force. Subsequently, the blue ball begins to move. It is clear that the red ball causes the blue ball to move in virtue of its momentum. The property *red*, on the other hand, is causally irrelevant.

In general, then, events involved in causal relations will maintain properties which are causally relevant and properties which are not. The goal is, first, to establish a set of criteria by which the causally relevant properties are clearly distinguished from the causally irrelevant, and second, to show that mental properties meet such criteria.

Meeting these goals will go a long way towards removing any remaining worries over the causal efficacy of the mental.

According to Fodor, in the antecedent event, what distinguishes the ball's being red from the ball's having a certain momentum (i.e. having a certain mass and moving at a certain rate of speed) is that the latter property is nomologically sufficient for the blue ball's subsequent movement, while the former is not. Ordinarily, any other ball with the same mass and speed, when it collides with a stationary blue ball, will produce the same subsequent movement. However, there are other red balls, with different masses and/or speeds, that are incapable of moving the blue ball. Since the ball's redness is not nomologically sufficient for the blue ball's movement, the redness is causally inert. On the other hand, since the red ball's mass and speed are nomologically sufficient for the blue ball's movement, they are causally efficacious. So causal efficacy can be understood in terms of causal law.

Not only does the present example illustrate that causal efficacy can be understood in terms of causal law, it also shows that events are subsumed under such causal laws in virtue of certain properties. The event of the red ball's colliding with the blue ball is sufficient for the event of the blue ball's motion in virtue of particular properties of the antecedent event. The relevant law would cover those antecedents which have the property of such-and-such mass and

speed. Thus, objects are subsumed by causal laws in virtue of certain properties "projected" by those laws.

Therefore, it is in virtue of a property (or properties) that an antecedent event is nomologically sufficient for its effect. As Fodor claims, "P is a causally responsible property if it's a property in virtue of which individuals are subsumed by causal laws." (Fodor, 143) So property P is causally responsible for the production of events with property E if events with property P are nomologically sufficient for events with property E.

So, as Fodor claims, "intentional properties are causally responsible in case there are intentional causal laws." (143) For purposes of justifying the causal-intentional relevance relation, which simply states that certain mental properties are causally relevant to certain behavioral events, it is enough to show that there are such intentional laws. If Fodor is right, the establishment of the existence of such laws, in addition to the work done in this thesis to allay any fears instilled by physicalism (such as systematic overdetermination), should be sufficient for the vindication of mental causation.

However, while it is generally maintained that intentional laws between mental events and behavioral events do exist, it is not clear that such laws are strong enough to support causal claims. The claim that an event can be regarded as a cause of a second event only if the former is

nomologically sufficient for the latter seems to imply that such events must be covered by strict laws. As Fodor claims,

> Causal transactions must be covered by exceptionless laws; the satisfaction of the antecedent of a covering law has to provide literally nomologically sufficient conditions for the satisfaction of its consequent so that its consequent is satisfied in every nomologically possible situation in which its antecedent is satisfied. (Fodor, 147)

It is doubtful, however, that mental events are covered by such exceptionless laws. Most philosophers of the mind (including Fodor) embrace the "anomalism" (or anomaly) of the mental: There are neither strict psycho-physical laws relating mental events and bodily events, nor strict psychological laws relating mental events to one another.

An example will illustrate the point. If I perform the behavior of walking to the refrigerator and getting a beer, it is reasonable to assume that I might very well have done so because I believed there was beer in the refrigerator and I desired to drink a beer. Thus, if it can be claimed that there are situations in which such a belief/desire pair successfully explains beer-retrieving behavior, it must also be claimed that the belief/desire pair regarding beer in the refrigerator is causally related to my subsequent behavior in virtue of being covered by a strict law: Such a belief/desire pair is nomologically sufficient for the production of this type of behavior.

The problem is, such a strict connection between such belief/desire pairs and subsequent behavior will jeopardize

our common conception of agency, viz. precluding deliberation, any notion of free will and the possibility of countervailing beliefs and desires. On such a conception of causality, behavior must follow, unfailingly, upon my belief and desire. This rules out the possibility that I might have beliefs and desires that overrule my belief that there is beer in the refrigerator and my desire for beer.

For instance, I might have a drinking problem. In such a situation, I will more than likely have the belief/desire pair regarding the contents of the refrigerator. However, one could easily imagine a situation in which I resist my desire for beer, due to countervailing beliefs and desires; namely, the belief that if I drink beer, I will fall back into patterns of binge drinking and self-destructive behavior, and the desire not to return to such patterns. It is quite conceivable that, due to this countervailing belief/desire pair, I will refrain from drinking the beer.

Not only are the capacity to deliberate and the possibility of countervailing beliefs and desires vital to our conception of agency, we encounter them in our daily lives. Sometimes, one might put off the possibility of immediate gratification for the sake of long-term gain. Other times, one's immediate beliefs and desires are victorious. I might believe that my favorite TV show is on and I might desire to watch it. However, I have the countervailing desire to finish my thesis, as well as the belief that if I watch my show, I won't do the work necessary

to complete my thesis. On any particular day, this desire may or may not prevail over my desire for immediate gratification.

Given such common experiences, along with the necessity of deliberation and countervailing beliefs and desires to our notion of agency, the claim that mental events (e.g. belief/desire pairs) and subsequent behavior are covered by strict laws is highly dubious. In other words, belief/desire pairs cannot be regarded as nomologically sufficient for subsequent behavioral events. But if causal relations must be backed by strict causal laws, then it appears that there is no basis to the claim that particular mental events are causally responsible for behavioral events. The project for Fodor is to make sense of the causal efficacy of the mental, relative to the anomaly of the mental and the standard requirement that causal claims be backed by strict causal laws.

Fodor's strategy is to claim that, while mental events and subsequent behavioral events are not covered by strict laws, they are covered by ceteris paribus laws (i.e. laws who's antecedent events have attached ceteris paribus clauses). According to Fodor's argument, ceteris paribus laws do the same work as strict laws in covering causal events: "If it's a law that M -> B ceteris paribus, then it follows that you get Bs whenever you get Ms *and the ceteris paribus conditions are satisfied.*" (Fodor, 152, his emphasis) In other words, where M and B are events covered by a ceteris

paribus law, as long as the ceteris paribus clause is fulfilled, M is sufficient for B.

Fodor's main point is that, if strict covering laws are sufficient for establishing a causal relation, then ceteris paribus laws will also be sufficient, since they both do the same work. Strict laws establish a causal relation, since the occurrence of the antecedent event is sufficient for the occurrence of the consequent event. But, in cases where the ceteris paribus clause of such a hedged law is discharged, the antecedent becomes sufficient for the consequent event. Thus, even if intentional laws are not strict and all we can ever expect from the mental are ceteris paribus laws, such laws establish causal relations just as well as strict laws. Hedged laws only require that the ceteris paribus conditions of that law be discharged for the antecedent to guarantee the consequent.

There is much to recommend Fodor's strategy. First, it appears that ceteris paribus clauses are not unique to psychology. All of the laws in the special sciences involve ceteris paribus clauses. Thus, if ceteris paribus laws do not back causal claims for psychology, then there is no reason to believe they do so for any special science.

Second, experience shows that ceteris paribus laws in psychology provide effective (although not flawless) predictions of human behavior: "We do use commonsense psychological generalizations to predict one another's

behavior; and the predications do--very often--come out true." (Fodor II, 4)

Not only are these predications reliable, they are often more accessible than those that the physical idiom is capable of providing. For instance, in order to predict where I will be next Saturday night, using only the language of physics, one would have to be Laplace's demon: There is far too much one needs to know in order to predict my whereabouts, using only the physical sciences. There is a far better, more useful strategy in predicting where I will be on Saturday night: As Fodor says, "far the best way to find out (usually, in practice, the *only* way to find out) is: *ask me!*" (Fodor II, 6, his emphasis).

Such a strategy requires that one invoke a ceteris paribus law: "If he says (believes) that he will be at the Green Door on Saturday night, then (ceteris paribus) he will be at the Green Door on Saturday night." Our best, and usually only, tools for predicting complex behavior are the ceteris paribus laws of the intentional idiom. According to Fodor's strategy, these laws back the causal claims made in psychology. And what better vindication of these ceteris paribus laws than their successful employment in practice.

Fodor's strategy for justifying the causal efficacy of the mental will not be defended here. Our present purpose is merely to offer strategies by which the causal-intentional relevance relation might be justified. Again, the causal-intentional relevance relation implies that the mental is

causally relevant when seeking an explanation of certain types of behavior (i.e. behavior subject to intentional description). If Fodor's strategy is successful, then the causal-intentional relation will be justified, in that causal-intentional laws will justify (certain) causal claims regarding mental states.

The most compelling doubts about the causal efficacy of the mind are raised by the implications of physicalism: How are we to understand the unique work done by the mental, given that the physical sciences are capable of generating exhaustive causal analyses of any behavioral event? How can one claim that the mental is both irreducible and causally efficacious and still avoid the problem of causal overdetermination? This thesis, I believe, presents a defensible solution to these problems. Further, Fodor offers a solid strategy for justifying the causal-intentional relevance relation. These arguments and strategies go a long way toward allowing us to understand and justify the claim that the mind is both irreducible and causally efficacious.

However, an important question remains: Might there not be a simpler solution to the problem of mental causation? Instead of making room for the causal efficacy of the mind, might we not instead look to the arguments of the eliminative materialist, who rejects the intentional idiom as a legitimate means of characterizing human cognition. If eliminativism is correct, then the problem of mental causation dissolves: Without mental states, there is no

problem of mental causation. The difficulties which motivate Kim's explanatory exclusion principle are avoided.

Thus, does eliminative materialism provide a simpler, more effective way to resolve the problem of mental causation than the one provided here? This question is the subject for the fifth and final chapter of this thesis. A response to the eliminativist position will not only demonstrate the value of the intentional idiom, but it will provide a final demonstration of the value of the pragmatic theory of explanation.

Chapter V: The Future of Intentionality

The foregoing considerations demonstrate that the pragmatic conception of explanation provides the basis for a plausible analysis of mental causation. Through the tools of this theory of explanation, we are able to make sense of and make room for the unique causal work performed by the mental in the production of behavior, without violating the central tenants of physicalism. But while this view is plausible, might there not be a better way to resolve the difficulties associated with mental causation?

It could be argued, for instance, that the contemporary problems with mental causation, such as causal overdetermination, are indicative of a deeper, more fundamental difficulty. Folk psychology (or common-sense intentionality) has been the central theory for predicting and explaining behavior for thousands of years. Talk of beliefs, desires, wants and wishes has been the primary way to understand ourselves as agents and to explain and predict the behavior of both strangers and close relations.

Intentionality has been so successful as a theory of human action that, as Fodor claims, it has become "practically invisible." (Fodor II, 3) Just as the expert tennis player, due to years of successful practice and use, no longer notices the distinction between herself and her tennis racket, we recognize no distinction between ourselves and the theory of mental states by which we explain and

predict human behavior. But to what do we attribute the success of the mental idiom? Do we credit the power and scope of folk psychology for its success or has it succeeded because it has long held a monopoly over characterizing, explaining and predicting human behavior?

While intentionality has enjoyed a great deal of success, it has also enjoyed a noticeable lack of competition. Slowly, that is beginning to change. Only recently have scientists begun to scratch the surface of the mysteries of the human brain. While neuroscience is a relatively new field of study, the prospects for the growth of its predictive and explanatory power are tremendous. If a fully articulated science of the brain were developed, would we still require a theory of the mental?

To put the matter another way, are the epistemological and metaphysical difficulties that motivate Kim's explanatory exclusion principle problems that require resolution, or are they rather an indication that intentionality has outlived its usefulness; that it must be eliminated as a legitimate characterization of human cognition and as a source for predicting and explaining behavior? Without a legitimate alternative, intentionality can survive and flourish. Problems such as causal overdetermination pose no real threat to the mental if there is no alternative. Now, there appears to be such an alternative, at least in principle.

Therefore, it could be argued that the pragmatic theory of explanation does not provide the best solution to the

problems associated with mental causation. Instead, the problems of mental causation can be dissolved by eliminating the mental as a legitimate means of characterizing human behavior and cognition. The contemporary difficulties of mental causation indicate that the mental has simply outlived its usefulness.

Thus, should philosophers of the mind embrace the eliminativist position? Is the elimination of the mental as a legitimate means of characterizing human behavior and cognition a superior solution to the problems of mental causation? In short, is intentionality scientific "dead wood"?

1. The Prospects for Eliminativism

Broadly speaking, the eliminativist believes that folk psychology is fundamentally wrong or misleading. Anti-reductionist attempts to retain an irreducible, useful theory of the mental are wasted philosophical effort. But why exactly is folk psychology wrong? Fodor, to take one example, is quick to point out that folk psychology has proven to be a useful theory for describing human cognition and predicting behavior (with astonishing accuracy) for thousands of years. What are the motivations behind eliminative materialism that could possibly justify eliminating folk psychology as a legitimate theory of human cognition and behavior?

Specifically, there are two types of eliminativism, each varying in degrees of strength. The first type we might call ***prescriptive eliminativism***. Representative of this view is Paul Feyerabend. Feyerabend's criticisms are directed at identity theory--the idea that mental states can be systematically reduced to physical states. Feyerabend claims that identity theory is a result of understandable but misguided presuppositions about how the development of a theory of the brain should proceed:

> Such [identity] hypotheses are usually put forth by physiologically inclined thinkers who want also to be empiricists. Being physiologically inclined, they want to assert the material character of mental processes. Being empiricists, they want their assertion to be a testable statement about mental processes. (Feyerabend, 266)

According to Feyerabend, for the identity theorist, the progress of a science of the brain will be gauged by how well it accounts for our mental lives. In other words, the goal of neuroscience is to generate a better, fuller understanding of the human mind. A fully developed and articulated theory of the brain will provide a deeper understanding of, for example, the fear of heights, addiction to drugs, the desire for chocolate, etc. Thus, while we satisfy our materialist inclinations by maintaining the physical character of the mental, we satisfy our "empiricist" inclinations by claiming that hypotheses about the brain should be "testable statements about mental processes."

Both our "physiological" and "empirical" inclinations are combined in an identity theory, where mental kinds are type-identical to physical kinds. A successful theory of the brain will allow us to develop the specific bridge laws between physical and mental states. But, according to Feyerabend, if we are to be materialists, such a theory is in principle incorrect.

To invoke an identity between mental states and physical states is to return to dualism: "[Identity theory] not only implies, as it is intended to imply, that mental events have physical features; it also seems to imply ... that some physical events ... have non-physical features. It thereby replaces a dualism of events with a dualism of features." (ibid, 266) Even with a type-identification between mental and physical states, there remains a dualism of properties or features. Thus, to embrace the identity theory is to embrace a sort of dualism. Therefore, since materialism is a monistic theory, identity theory is in principle incompatible with materialism, since it implies a dualism of properties or features.

Feyerabend claims that, if we are to be consistent materialists, a theory of the brain must be developed without regard to the theory of the mental. Whether we conceive of the mental as irreducible to the brain, or as type-reducible to brain states, the mental is *a priori* incompatible with our monistic metaphysics. Thus, to be consistent materialists,

we must give up talk of mental states altogether and develop a purely physiological conception of the human being.

Prescriptive eliminativism is an in "principle argument": We ought to give up talk of the mental, because such talk is a priori inconsistent with our materialist metaphysics. If Feyerabend's position turns out to be correct, then the problem of mental causation is dissolved. There is no problem of mental causation because a causal analysis of a behavioral event based in the mental idiom is fundamentally incorrect. Strictly speaking, there *is* no causal analysis involving mental states, since there are no mental states. In this way, Feyerabend's eliminativism provides an easy solution to the problem of mental causation.

A second form of eliminativism we might call ***predictive eliminativism***. Representative of this theory is Paul Churchland. Like Feyerabend, Churchland advocates the elimination of the mental as an adequate theory of human cognition. However, Churchland's arguments are not "in principle," a priori arguments. Instead, Churchland bases his eliminativism on the demonstrated inadequacies of folk-psychology and the historic parallels with other outdated, discarded theories.

The difference between Feyerabend's brand of eliminativism and that of Churchland can be demonstrated through their respective reactions to identity theory. Churchland, like Feyerabend, believes that a reduction of mental states to brain states is in principle impossible.

But while Feyerabend denies identity theory because of its dualistic implications, Churchland does so for different reasons:

> The one-to-one match-ups [between physical and mental types] will not be found, and our common-sense psychological framework will not enjoy an intertheoretic reduction, *because our common-sense psychological framework is a false and radically misleading conception of the causes of human behavior and the nature of cognitive activity*. (Paul Churchland, 43, his emphasis)

Churchland's brand of eliminativism proceeds from the problems and limitations of folk psychology. Folk psychology has survived, not because it is the *best* theory for characterizing human cognition and behavior, but because it has been the *only* theory. As Churchland claims, once a science of the brain has been fully developed, "we must expect that the older framework will simply be eliminated, rather than be reduced, by a matured neuroscience." (ibid. 43)

But what exactly are these supposed limitations of folk-psychology? First, Churchland cites what he calls the "explanatory poverty" of the intentional idiom. There are a number of capacities, requirements and characteristics of human beings that folk-psychology has not been able to explain successfully:

> We do not know what *sleep* is, or why we have to have it, despite spending a full third of our lives in that condition. ... We do not understand how *learning* transforms each of us from a gaping infant to a cunning adult, or how differences in *intelligence* are grounded.

> We have not the slightest idea how *memory* works, or how we manage to retrieve relevant bits of information instantly from the awesome mass we have stored. We do not know what *mental illness* is, nor how to cure it. (ibid. 46, his emphasis)

Fodor, on the other hand, repeatedly points out how well folk psychology has worked as a theory for predicting how others will behave (Fodor II, 3). Churchland admits that under limited conditions, folk psychology appears to give good behavioral predictions. But as soon as we venture beyond this limited scope (into the arena of patients who suffer from brain damage, for instance), folk psychology loses its explanatory and predictive power. It's predictive and explanatory scope are simply too narrow to be of any real value (Churchland, 46).

In addition to citing the supposed explanatory, predictive and descriptive poverty of folk psychology, Churchland draws parallels to the history of science, where the creation of a successful scientific theory led to the demise of a previous theory. For instance, Churchland cites the example of caloric, the theory of heat that held sway for much of 18th and 19th century physics. Heat was thought to be the substance caloric, a fluid contained in all objects. The transfer of heat from one body to another was thought to be a transfer of substance--a transfer of a certain quantity of caloric.

This theory enjoyed a degree of predictive and explanatory success, but was ultimately replaced by the theory of kinetic energy. Instead of heat being a *substance*

contained in an object, it had become accepted that heat is the *motion* of molecules of the object itself. This latter theory enjoyed greater explanatory and predictive success. Eventually, the theory of caloric was simply dropped by the scientific community.

As Churchland claims, it was not that the theory of caloric was incomplete. It was simply mistaken. Hence, it was eliminated as a theory of heat, even though it had enjoyed a limited degree of success. Similarly, Churchland predicts, once a more complete theory of the brain is developed, mental states, like caloric, will drop from our ontology.

Clearly, Feyerabend's version of eliminativism makes bolder claims than Churchland's. While the latter merely predicts the elimination of folk psychology, based on its demonstrated lack of explanatory and predictive scope and power, the former version eliminates the mental in principle, regardless of its value as a theory of human cognition and behavior. Let us begin by responding to Feyerabend's arguments.

Not only is Feyerabend's prescriptive eliminativism stronger, but it appears too strong to be plausible. Arguments similar to those used in chapter II to call into question the plausibility of the explanatory exclusion principle can be applied to Feyerabend's arguments: If Feyerabend's a priori arguments for the elimination of the mental are successful, then there seems to be no reason why

they won't be successful for other special sciences[1]. Sciences such as aerodynamics (with concepts such as the airfoil) invoke non-physical properties, i.e. properties not contained in the nomenclature of basic physics. So if we apply Feyerabend's arguments (and there appears to be no good reason why we should not), all such sciences and their ontologies must be eliminated. Surely this is absurd.

Thus, prescriptive eliminativism suffers from the same drawback as Kim's explanatory exclusion principle: As long as we apply them only to folk psychology, such theories appear plausible. But the scope of both Kim's and Feyerabend's arguments is too broad. Nothing precludes their application to *any* of the special sciences--sciences which repeatedly demonstrate their predictive and explanatory value. Once such an application is made, the untenability of their arguments is revealed.

The fact that prescriptive eliminativism becomes untenable when applied to less controversial special sciences provides a diagnosis of the problem with Feyerabend's argument: Arguments that, a priori, eliminate theories or explanations based solely on metaphysical concerns put the proverbial cart before the horse. One of the primary guides for theory retention or elimination is the explanatory and predictive power of a theory. Tyler Burge makes this type of

---

[1] See Chapter II, pp. 34-36.

argument regarding the value of the mental idiom in providing causal explanations:

> The probity of mentalistic causal explanation is deeper than the metaphysical considerations that call it into question. ... Materialist metaphysics is not the most plausible starting-point for reasoning about mind-body causation. Explanatory practice is. (Burge, 117-118)

Feyerabend's arguments for eliminativism fail because they presuppose that the value of a scientific theory is determined primarily by how well it fits into a monist metaphysics. If a theory does not fit neatly into such a metaphysics, it must be eliminated, regardless of its practical value.

When we apply his arguments to other special sciences, those arguments lose their luster because it is evident that such theories are valuable, not because they fit nicely into a materialist metaphysics (for if Feyerabend is correct about the incompatibility of property dualism and substance monism, then surely they do not), but because they have considerable predictive and explanatory power. Further, they can provide explanations that the physical sciences cannot. Thus, we can broaden Burge's claim: The probity of the explanations (and explanatory value) of any special science is deeper than the metaphysical considerations that call them into question.

That is not to say that metaphysical considerations are completely removed from theory evaluation. One ought not think that the door is open to whatever occult property one wishes. We can consider our metaphysical commitments as one

set of criteria for theory evaluation, but we cannot see them as the primary or sole criteria. Feyerabend's arguments make such an assumption. Subsequently, his arguments are rendered untenable.

On the other hand, Churchland's *predictive eliminativism* is a much more plausible argument. On his argument, the mental idiom will be eliminated because of its supposed explanatory and predictive poverty. Because it has failed to characterize much of human cognition (such as learning, memory, imagination, etc.) it will be eliminated in favor of a developed theory of the brain which, it is expected, will not suffer the same explanatory and predictive limitations.

Churchland's arguments avoid the problems encountered by Feyerabend's, since there are no implications for eliminating other special sciences. On predictive eliminativism, theories can be evaluated on an individual basis according to their descriptive, explanatory and predictive value, relative to the prospects of a (potential) replacing theory or science. Thus, predictive eliminativism makes a weaker but more tenable argument than prescriptive eliminativism.

But while predictive eliminativism cannot be ruled out in principle, it is highly doubtful that Churchland's predictions will come to fruition. First, the bulk of this thesis has been designed to show that the mental idiom is capable of generating causal explanations of behavior that the physical idiom cannot. Further, not only are these causal explanations unique to the mental, they are vital to

our conception of ourselves as rational agents. Tyler Burge defends the uniqueness and importance of the intentional idiom via its ability to generate unique causal explanations of behavior:

> Much of the interest of psychological explanation, both in psychology and in ordinary discourse, lies in helping us understand ourselves as agents. ... We think that we make things happen because we make decisions or will to do things. We think that we make assertions, form theories, and create cultures, because we think certain thoughts and have certain goals--and we express and fulfill them. In this context, we identify ourselves primarily in terms of our intentional mental aspects ... Our agency consists in our wants', willings', thoughts', values as such ... having some sort of efficacy in the world. Our mental events having the intentional characters that they have is, in individual instances, what we define our agency in terms of. (Burge, 118-119)

The intentional idiom provides a unique way of viewing human behavior and the payoff of this mode of inquiry is in the unique explanations and characterizations of human behavior that it provides. It is primarily through the intentional idiom that we understand ourselves as rational agents.

Such an understanding is not captured in the physical idiom. Rationality is unique to the mental and is not to be found in the physical idiom. For instance, arguments have been put forth (primarily by Donald Davidson) to suggest that mental states cannot be reduced to physical states *simply because* the rationality of beliefs and actions is not captured by reference to physical states. Rationality is unique to the intentional idiom and cannot be assimilated by

the physical. Cynthia and Graham Macdonald offer examples of this "rationality of beliefs":

> Mental states, such as belief states, have normative connections with each other: if one believes that 2+2=4, and that 1+1+1+1=4, then one *ought* to believe that 2+2=1+1+1+1. If one wants to be a good chess player and believes that studying chess openings will improve one's game, then one ought to study chess openings. This would be rational, given this belief and desire. ... The normativity involved in these connections is essential to the mental and is not to be found in the physical domain. No purely physical property has *rational* connections to other physical properties, so a reduction is in principle impossible. (Macdonald, 9)

It is the intentional aspect of the mental that allows for the characterization and explanation of human behavior as purposive and rational. While such a characterization is unavailable to the physical idiom, it is vital to our common conception of ourselves as agents[2].

Thus, the comparison Churchland draws to historical examples is not as persuasive as one might imagine. For instance, the theory of caloric makes a poor comparison to the mental idiom because it is difficult to find the explanatory or predictive power that is unique to that theory. In other words, it is not clear that caloric, as a theory of heat, does any explanatory or predictive work that is not accomplished by the theory of kinetic energy. If this

---

[2] It can be argued by extension that intentionality is necessary to a successful analysis of free will. See, for instance, C.A. Campbell's notion of "phenomenological analysis" in "Has the Self a 'Free Will'?" in On Selfhood and Godhood (London: George Allen & Unwin) 1957, pp. 158-179.

thesis is correct and the intentional idiom is capable of generating unique and important causal explanations of behavior, then comparisons to examples such as caloric simply do not hold.

Therefore, while Churchland's arguments are not in principle untenable, it is doubtful that his predictions regarding the eventual elimination of the mental idiom will come to fruition. If, indeed, a mature neuroscience were to capture both the unique character of human cognition and the explanations of human behavior found in intentionality, then Churchland's predictions might very well be realized. However, for such a study of the brain, this would require a radical break from the physical sciences, given that rationality and agency are not concepts currently contained in such sciences.

Further, Churchland is correct in claiming that the explanatory and predictive scope of a mature neuroscience will be much broader than that of folk psychology. This fact is acknowledged in chapter IV: The e-contexts of the mental idiom can suffer from *descriptive failure*, while in principle, the physical idiom cannot. The physical idiom, it is granted, is capable of generating a true explanation of any behavioral event, while the intentional idiom is not. On the other hand, while the descriptive, explanatory and predictive capabilities of the mental idiom are narrow, they are nonetheless valuable. They are vital to our conception of ourselves as rational agents and they cannot be captured

in the physical idiom. Thus, regardless of the narrow predictive and explanatory scope of folk psychology, it is doubtful that it will simply be eliminated in favor of a fully articulated neuroscience.

Given our analysis of and response to both prescriptive and predictive eliminativism, we are now in a position to answer the central question of this chapter; namely, can eliminativism offer a superior resolution to the problems of mental causation? While eliminativism is capable of generating a plausible solution to the problem of mental causation, that solution is by no means superior. The cost of embracing that solution is our common conception of ourselves as rational agents and that cost, most would agree, is much too high.

On the other hand, the solutions offered in this thesis resolve the difficulties of mental causation. Further, they retain the important, albeit limited, explanatory, descriptive and predictive work performed by the intentional idiom. The fact that, with the pragmatic conception of explanation, we are able to resolve the difficulties of mental causation and still retain the valuable work performed by the intentional idiom makes the solution offered in this thesis superior to the solution offered by the eliminativist.

2. Concluding Remarks: The Value of the Pragmatic Theory of Explanation

If anti-reductionist arguments are correct, then mental states will resist a systematic reduction to physical states. But while a reduction of the mind to the body may be avoidable, an assimilation of the mind into our scientific world-view is a necessity. A modern conception of the mind must find its place in a physical world.

While there are a number of difficulties and obstacles to such an assimilation, none are more obstinate and important than the problems of mental causation: Given that any behavioral event can be given an exhaustive causal analysis in the physical idiom, what work is left for the mental to do? Further, if there is a sense in which the mental does unique causal work in the production of behavior, how are we to avoid the problem of causal overdetermination?

The pragmatic conception of explanation provides the foundation for an analysis of mental causation that answers both of these questions. Clearly, however, this theory of explanation is still in its infancy, despite the work of theorists such as van Fraassen and Garfinkel. It will need to be refined in order to deal with the inevitable objections and counterexamples faced by any theory of explanation. But while it requires further development and refinement, there is much to recommend this pragmatic theory of explanation.

First, the pragmatic theory of explanation is valuable because the analysis of mental causation that it provides is

valuable. Such a theory of explanation provides an understanding of mental causation that makes sense of and makes room for the unique causal powers of the mental. The causal efficacy of the mental, in turn, gives value and meaning to human agency. As Burge argues,

> If intentional psychological explanation 'made sense' of what we did ... but did not provide insight into the nature of any causal efficacy, it would lose much of its point. It would provide no insight into the various forms of agency that give life its meaning and purpose, and psychology its special interest. (Burge, 119)

If we are unable to answer the problems of mental causation, then we must either embrace epiphenomenalism, or eliminate the mental altogether. As has been argued in this thesis, both of these latter options come at too high a price. Hence, regardless of the difficulties that the pragmatic theory might face, it demonstrates its value and usefulness in the understanding of mental causation that it generates.

Further, the pragmatic theory of explanation can avoid the explanatory exclusion principle, but it is not clear that other theories of explanation are so able. Recall that the explanatory exclusion principle requires that there be only one complete and independent explanation for any event or phenomenon. As shown in chapter I, Kim offers metaphysical and epistemological arguments that motivate the explanatory exclusion principle, effectively precluding the possibility of having multiple complete and independent explanations.

However, it has been shown that the explanatory exclusion too strong a condition on explanation in general. The implications of the explanatory exclusion principle are such that the irreducible properties of any special science must be regarded as causally inert. Thus, instead of being a solution to the problem of mental causation, it is a problem for explanation theory: In order to avoid the implications of the explanatory exclusion principle, a theory of explanation must be able to resolve the metaphysical and epistemological problems of having (apparently) multiple complete and explanations. In other words, a satisfactory theory of explanation must be able to cope with the difficulties of compounding epistemic puzzlement and causal overdetermination.

As shown in chapter III, the pragmatic theory is capable of diffusing these difficulties--it is able to maintain a multiplicity of explanations without compounding puzzlement and without implying causal overdetermination. On the other hand, it is not clear that other standard theories of explanation, such as Hempel's D-N model, are able to deal with these epistemological and metaphysical difficulties. Therefore, regardless of the difficulties or objections that it may encounter, the pragmatic theory of explanation shows its value in that it is able to avoid the difficulties which entail explanatory exclusion.

Churchland may be right. The intentional idiom may eventually find itself on the scientific scrap heap. But

there are good reasons for believing that Churchland is wrong--that the mind and its unique causal powers are not only assimilable into a scientific world-view, but also that they are vital to our conception of ourselves as agents. A successful analysis of mental causation will maintain the causal efficacy of the human agent in the world and help assimilate the mind and its unique causal properties into our scientific world-view. I believe this thesis has provided such an analysis.

BIBLIOGRAPHY

BIBLIOGRAPHY

Achinstein, Peter .The Nature of Explanation. Oxford: Oxford University Press, 1983.

Barnes, Jonathan. Early Greek Philosophy. London: Penguin Books, 1987.

Batens, Diderik. "Do We Need a Heirarchical Model of Science?" Inference, Explanation, and other Frustrations: Essays in the Philosophy of Science. Ed. John Earman. Berkeley: University of California Press, 1992. 199-215.

Bromberger, Sylvian. On What We Know We Don't Know: Explanation, Theory, Linguistics and How Questions Shape Them. Chicago: The University of Chicago Press, 1992.

Burge, Tyler. "Mind-Body Causation and Explanatory Practice." Mental Causation. Ed. John Heil and Alfred Mele. Oxford: Clarendon Press, 1995. 97-120.

Burtt, E.A.. The Metaphysical Foundations of Modern Science. Atlantic Highlands, NJ: Humanities Press, 1952.

Campbell, C.A.. On Selfhood and Godhood. London: George Allen & Unwin, 1957.

Carnap, Rudolf. "Empiricism, Semantics and Ontology." The Philosophy of Science. Ed. Richard Boyd, et al. Cambridge, Mass: The MIT Press, 1991. 85-98.

Churchland, Patricia S. Neurophilosophy. Cambridge, Mass.: The MIT Press, 1993.

Churchland, Paul. Matter and Consciousness. Cambridge, Mass: The MIT Press, 1993.

---. "Eliminative Materialism and Propositional Attitudes." Journal of Philosophy. 78, 1981. p. 73.

Davidson, Donald. "Thinking Causes." Mental Causation. Ed. John Heil and Alfred Mele. Oxford: Clarendon Press, 1995. 3-18.

---. "Mental Events." Essays on Actions and Events. Oxford: Clarendon Press, 1980. 207-224.

---. Inquiries into Truth and Interpretation. Oxford: Clarendon Press, 1984.

Dretske, Fred. "Reply: Causal Relevance and Explanatory Exclusion." Philosophy of Psychology: Debates on Psychological Explanation. Ed. Cynthia MacDonald and Graham MacDonald. Oxford: Blackwell Publishing, 1995. 142-155.

Feyerabend, Paul. "Mental Events and Brain Events." The Nature of Mind. Ed. David M. Rosenthal. Oxford: Oxford University Press, 1991. 266-267.

Feynman, Richard P. QED: The Strange Theory of Light and Matter. Princeton, NJ: Princeton Science Press, 1985.

Fodor, Jerry. "Making Mind Matter More." A Theory of Content and Other Essays. Cambridge, Mass.: The MIT Press, 1992. 137-160.

---. Psychosemantics. Cambridge, Mass.: The MIT Press, 1993.

---. "You Can Fool Some of The People All of The Time, Everything Else Being Equal; Hedged laws and Psychological Explanation" Mind 100. (1991): 19-33.

Friedman, Micheal. "Explanation and Scientific Understanding." Theories of Explanation. Ed. Joseph Pitt. New York: Oxford University Press, 1988. 188-198.

Gamow, George. The Great Physicists From Galileo to Einstein. New York: Dover Publications, 1961.

Garfinkel, Alan. Forms of Explanation: Rethinking The Questions of Social Theory. New Haven: Yale University Press, 1981.

Gasper, Philip. "Causation and Explanation: Introductory Essay." The Philosophy of Science. Ed. Richard Boyd, et al. Cambridge, Mass: The MIT Press, 1991. 289-298.

Hempel, Carl G. Aspects of Scientific Explanation and Other Essays in the Philosophy of Science. New York: The Free Press, 1965.

Hintikka, Jaakko. "Semantics and Pragmatics for Why-Questions." The Journal of Philosophy. 1995. pp. 636-657.

Holcomb, Harmon. "Logicism and Achinstein's Pragmatic Theory of Scientific Explanation." Dialectica 41, no. 3. (1987) 239-248.

Kim, Jaegwon. Supervenience and Mind: Selected Philosophical Essays. Cambridge: Cambridge University Press, 1993.

---. Philosophy of Mind. Bolder: Westview Press, 1996.

---. "Explanatory Exclusion and the Problem of Mental Causation." Philosophy of Psychology: Debates on Psychological Explanation. Ed. Cynthia MacDonald and Graham MacDonald. Oxford: Blackwell Publishing, 1995. 121-141.

---. "Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism?" Mental Causation. Ed. John Heil and Alfred Mele. Oxford: Clarendon Press, 1995. 19-26.

Kuhn, Thomas. The Structure of Scientific Revolutions. Chicago: The University of Chicago Press, 1962.

Laudan, Larry. Science and Values: An Essay on the Aims of Science and Their Role in Scientific Debate. Berekely: University of California Press, 1984.

MacDonald, Cynthia and Graham MacDonald. "Introduction: Supervenient Causation". Philosophy of Psychology: Debates on Psychological Explanation. Ed. Cynthia MacDonald and Graham MacDonald. Oxford: Blackwell Publishing, 1995. 4-28.

Mackie, John L. "Causes and Conditions." Causation. Ed. Ernest Sosa and Michael Tooley. Oxford: Oxford University Press, 1993. 33-55.

Menzies, Peter. "Against Causal Reductionism" Mind 97. (1988): 551-574.

Nagel, Ernest. Teleology Revisited and Other Essays in the Philosophy and History of Science. New York: Columbia University Press, 1979.

Priest, Stephen. Theories of the Mind. Boston: Houghton Miflin Company, 1991.

Putnam, Hilary. "The Nature of Mental States." The Nature of Mind. Ed. David M. Rosenthal. Oxford: Oxford University Press, 1991. 197-203.

Quine, W.V.O.. Pursuit of Truth. Cambridge: Harvard University Press, 1992.

Rorty, Richard. Consequences of Pragmatism. Minneapolis: University of Minnesota Press, 1982.

Rosenthal, David M. ed. Materialism and the Mind-Body Problem. Indianapolis: Hacket Publishing, 1987.

Salmon, Wesley C. "Four Decades of Scientific Explanation." Scientific Explanation. ed. Philip Kitcher and Wesley C. Salmon. Minnesota Studies in the Philosophy of Science, Vol. XIII. Minneapolis: University of Minnesota Press, 1989.

Scriven, Micheal. "Causation as Explanation." Nous 9. (1975): 3-16.

Sosa, Ernest. "Mind-Body Interaction and Supervenient Causation" Midwest Studies in Philosophy IX, (1984): 271-281.

Stillings, Neil A., et al. Cognitive Science: An Introduction. Cambridge, Mass.: The MIT Press, 1995.

Van Fraassen, Bas C. "The Pragmatic Theory of Explanation." Theories of Explanation. Ed. Joseph Pitt. New York: Oxford University Press, 1988. 136-155.

---. The Scientific Image. Oxford: Clarendon Press, 1980.

Wittgenstein, Ludwig. On Certainty. Ed. G.E.M. Anscombe and G.H. von Wright. Trans. Denis Paul and G.E.M. Anscombe. New York: Harper and Row, 1969.

Worley, Sara. "Mental Causation and Explanatory Exclusion." Erkenntnis 39. (1993): 333-358.