Southern Illinois University Carbondale **OpenSIUC**

Dissertations Theses and Dissertations

5-1-2010

The Tale of Mental Causation: Fact or Fiction?

Chia-Lin Tu

Southern Illinois University Carbondale, chialintu@gmail.com

Follow this and additional works at: http://opensiuc.lib.siu.edu/dissertations

Recommended Citation

Tu, Chia-Lin, "The Tale of Mental Causation: Fact or Fiction?" (2010). Dissertations. Paper 125.

This Open Access Dissertation is brought to you for free and open access by the Theses and Dissertations at OpenSIUC. It has been accepted for inclusion in Dissertations by an authorized administrator of OpenSIUC. For more information, please contact opensiuc@lib.siu.edu.

THE TALE OF MENTAL CAUSATION: FACT OR FICTION?

by

Chia-Lin Tu

B.A., Soochow University, Taiwan, 1996 M.A., National Chung-Cheng University, Taiwan, 1999 M.A., University at Albany–SUNY, 2005

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

> Department of Philosophy in the Graduate School Southern Illinois University Carbondale May 2010

Copyright by Chia-Lin Tu, 2010 All Rights Reserved

DISSERTATION APPROVAL

THE TALE OF MENTAL CAUSATION: FACT OR FICTION?

By

Chia-Lin Tu

A Dissertation Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in the field of Philosophy

Approved by:

Dr. Pat Manfredi, Chair

Dr. Douglas Anderson

Dr. Larry Hickman

Dr. Yu-wei Wang

Dr. Andrew Youpa

Graduate School Southern Illinois University Carbondale April 7, 2010

AN ABSTRACT OF THE DISSERTATION OF

Chia-Lin Tu, for the Doctor of Philosophy degree in Philosophy, presented on April 6, 2010, at Southern Illinois University Carbondale

TITLE: THE TALE OF MENTAL CAUSATION: FACT OR FICTION?

MAJOR PROFESSOR: Dr. Pat Manfredi

Mental causation is with us all the time. Being a table is different from being a human---although we are composed of physical particles, we have understanding, reason, or perception, which are able to make a difference in the physical world.

In this dissertation, I have detail discussions of contemporary substance dualism, the mind-brain identity theory, and Jaegwon Kim's functionalism, and thus conclude that none of them can provide an appropriate account to the problem of mental causation. By distinguishing the mind from the body, substance dualists face the pairing problem: How does this particular mind unite with this particular body and thus interact? With the pairing problem, more and more philosophers accept physicalism. However, it is surprising that the problem of mental causation arises again from the heart of physicalism. It means that accepting physicalist ontology does not make this problem go away. On the contrary, basic physical assumptions can even be seen as the source of the current difficulties with mental causation.

My preferred idea is that mental properties emerge from physical properties, and both of them together make an occurrence to cause an effect. Emergence makes mental causation autonomous and also avoids epiphenomenalism.

i

ACKNOWLEDGEMENT

I think I have been a graduate student too long. It was long enough that I forgot to see the end. But I did it!

First and foremost, I would like to express the deepest appreciation to my committee chair, Professor Pat Manfredi, who has character and heart of a respectable faculty. He conditionally and convincingly conveyed a spirit of adventure to research. Without his guidance and persistent help, this dissertation would not have been possible.

I would like to thank my committee members, Professor Douglas Anderson,
Professor Larry Hickman, Professor Yu-wei Wang, and Professor Andrew Youpa for
demonstrating concern and comments on my research and their patience with a slow
writer.

A special thank goes to Professor Douglas Berger, who as a big brother is always willing to provide best suggestions and to ease my stress.

For all friends who enter my life to make it more colorful, I thank you for tolerating my impatience. You always support me and encourage me with you best wishes.

Finally, I have to dedicate this dissertation to my family, especially my dear grandfather. I want to let him know that I have accomplished the promise we made a long time ago. And this dissertation would not be successful without my parents, whose unconditional love was my foundation during the time when I was alone studying in the States. They were always supporting me whatever I decided to do.

There is no other word but I love you all.

TABLE OF CONTENTS

ABSTRA	ACT	i
ACKNO	WLEDGEMENT	ii
LIST OF	F FIGURES	v
CHAPTER 1 Introduction		1
CHAPTER 2 Substance Dualism		10
2.1	Substance Dualism and the Pairing Problem	10
2.2	John Foster	15
2.3	Peter Unger	22
2.4	W. D. Hart	26
2.5	Conclusion: the Problem of Mental Causation Unsolved	29
CHAPT	ER 3 The Mind-Body Identity Theory	33
3.1	The Identity Theory The Old School	33
3.2	Philosophical Arguments against the Identity Theory	40
3.3	Christopher Hill, Brian McLaughlin and New Wave Materialism	43
3.4	Arguments for Mind-Brain Identities	48
3.5	Arguments against Mind-Brain Identities	54
CHAPT	ER 4 Functional Reduction	58
4.1	Varieties of Functionalism	58
4.2	Property Dualism, Reduction and Reductive Explanation	63
4.3	Reductive Explanations via Functional/Causal Role	66
4.4	The Limits of Kim's Model of Functionalism	70
4.5	Another Reductionist ModelBridge-Law Reduction	75

CHAPT	ER 5 Emergentism	80
5.1	Non-Reductive Physicalism	80
5.2	Emergent Properties	84
5.3	Mental Causation in a Physical World	89
5.4	A New Concept of Emergence	93
5.5	Defining Mental Causation within New Emergent Framework	97
References		
VITA		107

LIST OF FIGURES

Figure 1: New emergentist theory of menta	l causation	100
---	-------------	-----

CHAPTER 1

INTRODUCTION

The idea that we are agents who act for reasons is central to our self-conception. Imagine yourself in a situation in which you see a speeding bus coming towards you. Would you move to avoid the bus or remain still allowing it to hit you? It is safe to say that you would move because you do not want to be injured or die. This paradigmatic example explains why an action occurred by revealing a person's reasons for performing the action. What makes such an explanation true; that is, by virtue of what connection between a reason and an action does that reason explain the action?

Within analytic philosophy, there have been two approaches for answering this question: a causal approach¹ and a non-causal approach.² For the purpose of this dissertation, I am assuming that the causal approach is the correct one. This perspective lets us see the causal analysis of intentional behavior as part of the larger project of explaining the role of intentionality generally in our account of the world. Donald Davidson has made the causal power of reasons central to his analysis of actions, arguing that agents act for reasons only when their reasons cause their behaviors. More precisely, actions are intentional only if the actions are caused by the beliefs, desires, or other propositional attitudes of the agent who acts. He treats the connection between reasons and actions as a causal explanation linking between the two events. He says,

¹ Some examples were G. E. M. Anscombe, H. L. A. Hart, A. M. Honoré, and William Dray. See Donald Davidson's "Action, Reasons, and Causes," footnote 1.

² Some examples were Anthony Kenny and A. I. Melden. See Donald Davidson's "Action, Reasons, and Causes," footnote 1.

We cannot explain why someone did what he did simply by saying the particular action appealed to him; we must indicate what it was about the action that appealed.³

Whenever someone does something for a reason, the reason can be characterized as both a desire and a belief. Giving the reason why someone does something is a form of causal explanation. As Davidson himself puts it, in giving the "the agent's reason for doing what he did," we point to "some feature, consequence, or aspect of the action the agent wanted, desired, held dear, thought dutiful, beneficial, obligatory, or agreeable." In other words, it is natural to think that these features, consequences, or aspects of the actions were the agent's reasons for acting. The explanation of actions refers to an agent's intentions or motives in acting, *the* reason. The reason for an action is its cause. Thus, as I see it, our agency depends on the fact that our reasons causally bring about our actions.⁵

Generally, I agree with Davidson that the connection between an agent's reasons and actions performed for these reasons requires an explanation. The connection between reasons and actions is made by causal laws. On this approach, to say that someone has a reason for acting would be to say that anyone with such a reason would, *everything else being equal and ignoring any fundamental indeterminacy*, perform the same act. If the agent merely has a reason but did not act because of it, then there was another causal law applicable to the agent's action. The concept of reasons explicated in terms of causal laws

³ Donald Davidson, "Actions, Reasons, and Causes" The Journal of Philosophy Vol. LX, No. 23 (1963): 685.

⁴ Donald Davidson, "Actions, Reasons, and Causes" The Journal of Philosophy Vol. LX, No. 23 (1963): 685.

⁵ To claim that reasons cause actions does not entail any particular position with respect to free will. In this dissertation, I make no claim about whether we have free will or lack it. Moreover, assuming we have free will I make no claim about whether the causal connections between our reasons and actions are deterministic or indeterministic ones.

entails that intentional actions actually are governed by reasons. Most of the arguments over the causal approach have centered on whether there are such laws. Its advocates say that there are, whereas its critics say that there are not. Davidson sided with the critics by arguing that most people missed the point in thinking that it mattered whether or not there are causal laws connecting reasons and actions. Even if there are no causal laws between reasons and actions, there may be purely physical laws that explain the connection between reasons and actions. So if reasons cause actions, both the reasons and the action can be re-described in a physical vocabulary that uncovers a lawful connection between them. In this dissertation, I shall neither dispute what Davidson argued about causality nor his claim that the absence of causal laws is compatible with reasons causing behavior. While recent causal analyses of intentional action often differ from Davidson's in significant ways, all are indebted to his work. Thus, anyone who adopts this approach has been led to think of reasons for action as mental states that cause actions. An explanation of our agency presupposes some solution to the problem of mental causation. I assume the causal approach is correct and in these chapters, I will join the vigorous debate about the exact account to be given of mental causation.

In addition to adopting a causal approach to the connection between reasons and actions in this dissertation, I am adopting what might be called a "substance" conception of mentality rather than the purely "relationship" approach to mentality championed by William James and many of his pragmatist followers. James says,

If the reader will take his own experiences, he will see what I mean. Let him begin with a perceptual experience, the 'presentation,' so called, of a physical object, his actual field of vision, the room he sits in, with the book he is reading as its center; and let him for the present treat this complex object in the common-sense way as being 'really' what it seems to be, namely, a collection of physical things cut out from an environing

world of other physical things with which these physical things have actual or potential relations. Now at the same it is just *those self-same things* which his mind, as we say, perceives; and the whole philosophy of perception from Democritus's time downwards has been just one long wrangle over the paradox that what is evidently one reality should be in two places at once, both in outer space and in a person's mind. 'Representative' theories of perception avoid the logical paradox, but on the other hand they violate the reader's sense of life, which knows no intervening mental image but seems to see the room and the book immediately just as they physically exist.

The puzzle of how the one identical room can be in two places is at bottom just the puzzle of how one identical point can be on two lines. It can, if it be situated at their intersection; and similarly, if the 'pure experience' of the room were a place of intersection of two processes, which connected it with different groups of associates respectively, it could be counted twice over, as belonging to either group, and spoken of loosely as existing in two places, although it would remain all the time a numerically single thing.⁶

It is fundamental to James' theory of mind that there is a certain kind of parallelism between bodily processes and the stream of consciousness. In other words, what makes a particular event E a mental event is its membership in a continuous stream of events that together constitute one's personal consciousness. The very same event, E, by virtue of its relations to a physical object, is a physical event. E itself is neither mental nor physical. It is simply an event in pure experience. What makes E either mental or physical, is nothing more nor less than its *relations* to other events with which it is being conceived at a given time for some specific purpose.

James' account of mentality allows for a theory of mental causation that is sympathetic to what James regard as essential to the religious point of view. The relationship between the mental and the physical does not need to be mechanical. It could

_

⁶ William James, "Does 'Consciousness' Exist?" The Journal of Philosophy Vol. 1, No. 18 (1904): 481.

rather be a matter of design that the stream of consciousness and brain processes harmonize as they do. James' discussion of the causal relationship between the mental and the physical accords with the psychologists' principle, *No psychosis without neurosis*. He says,

The feelings can produce nothing absolutely new, they can only reinforce and inhibit reflex currents which already exist, and the original organization of these by physiological forces must always be the groundwork of the psychological scheme.⁷

The brain, for James, is basically a reflex organ, and a human being is a reflex organism. If thought influences actions, it does so through the inhibition or reinforcement of reflex arcs that has already traversed their afferent beginnings. That mental causation, for James, occurs as part of a conception of the universe as purposive, is suggested at the outset of *the Principles* in this remark,

Just so we form our decision upon the deepest of all philosophic problems: Is the Kosmos an expression of intelligence rational in its inward nature, or a brute external fact pure and simple? If we find ourselves, in contemplating it, unable to banish the impression that it is a realm of final purposes, that it exists for the sake of something, we place intelligence at the heart of it and have a religion. If, on the contrary, in surveying its irremediable flux, we can think of the present only as so much mere mechanical sprouting from the past, occurring with no reference to the future, we are atheists and materialists.⁸

James believed that the interaction between the mental and the physical, the pure experience taken twice over, has to be seen as an expression of purpose, a fundamental harmony between the stream of consciousness and brain processes. The pure experience, as I mentioned before, is neither mental nor physical, but it is purposive. Part of what this

5

⁷ William James, *The Principles of Psychology*. (Harvard University Press, 1983), 141.

⁸ William James, *The Principles of Psychology*. (Harvard University Press, 1983), 21.

means is that we can refuse the demand for a 'mechanism' that makes the causal connection between mental causes and their physical effects intelligible. With this account, James argued that speaking of mental causation is always speaking relative to any inquiry into a purpose. Mentality is not a 'thing' or 'substance' with causal power of its own, independent from a context of inquiry and explanation.

Despite its popularity among pragmatists, James' relational conception of mentality has had few adherents within the analytic tradition. Most philosophers in this tradition regard mental events as occurrences that take place within persons, where persons are, in turn, regarded as individual substances. This perspective is the one I am adopting. Of course, what makes such occurrences *mental* rather than *physical* has been a matter of considerable debate. Substance dualists believe the fact that such occurrences take place in a non-physical substance makes them mental. Many philosophers who reject substance dualism believe that the distinctive properties of such events, their intentional or qualitative content, are what make them mental. In either case, mentality is not a matter of the relationship between events belonging to a continuous stream of consciousness, but rather a matter of the distinctive properties possessed by mental events.

Returning now to the topic of mental causation, Jaegwon Kim agrees with me about its importance for our self conception. He says,

the objects around us to be rearranged. That is how we manage to navigate around the objects in our surroundings, find food and shelter, build bridges and cities, and destroy the rain forests.⁹

Kim not only thinks that mental causation is essential to our agency; he also believes that it is essential to our ability to think and speak as well.

...there seem to be compelling reasons for thinking that our capacity to think about and refer to things and phenomena of the world---that is, our capacity for intentionality and speech---depends on our being, or having been, in appropriate cognitive relations with things outside us, and that these cognitive relations essentially involve causal relations.¹⁰

Kim thinks that these are the reasons to save mental causation. Our moral practices such as beliefs, desires, intentions, and decisions, require our mentality which does have causal efficacy to our behaviors. Furthermore, when we develop our knowledge, we need our ability to speak and think which require perception, and thus we create beliefs about objects in this world. This involves a special relationship between what happens in our mind and what is real in the world. If we do not have perception, it seems that the foundation for most of our knowledge has been taken away. For Kim, the failure to provide an account of mental causation would undercut not only the conception of ourselves as practical and moral agents, but also the conception of ourselves as individuals who obtain knowledge about the world and who speak and think about things in that world.

Kim is not the first, nor the only philosopher to recognize the importance of mental causation. René Descartes struggled with the problem of mind-body interaction when he developed his revolutionary metaphysical picture in the 17th century. In chapter

⁹ Jaegwon Kim, *Physicalism, or Something Near Enough*. (Princeton University Press, 2005), 9.

¹⁰ Jaegwon Kim, *Physicalism*, or *Something Near Enough*. (Princeton University Press, 2005), 10.

two, we will examine what Kim calls "the pairing problem," a problem that he thinks dooms any substance dualist account of mental causation. We will also explore the attempts of three contemporary substance dualists to overcome this problem and to provide an account of mental causation.

In chapter three, we will examine the mind-brain identity theory and its consequences for the problem of mental causation. Why cannot mental processes just turn out to be brain processes; just as light turns out to be electromagnetic radiation, and the gene turns out to be the DNA molecule? If this were so, the problem of mental causation would be solved. Mental causation would simply be physical causation. We will see that although the identity theory provides an account of mental causation, it must be rejected for other reasons. We will explore Saul Kripke's reasons for rejecting the identity theory and discuss two attempts by "new wave materialists" to revive the identity theory despite Kripke's objection to it. Finally, I argue that despite these attempts to revive the identity theory, "the multiple realization argument" shows that mental properties are not reducible to physical properties. While mind-brain identity theory fails because of the multiple realization argument, that argument provided the motivation for an attractive alternative theory, namely functionalism. The functional perspective provides a new way to understand the correlation between the mental and the physical yet retain the autonomy of the mental. By distinguishing ontological reduction from reductive explanation, functionalism makes it possible for us the shed the constraints of the mind-brain identity theory without returning to substance dualism. Because functionalism preserves the distinction between mental properties and physical properties, it reopens the problem of mental causation. Is functionalism the remedy for the mindbody problem? Kim thinks that functionalism provides the best account of mental causation, but that it does not tell the whole story about mind and body. I disagree. I argue that Kim's account fails to explain adequately the causal power of mental properties. We will explore the whole picture associated with Kim's functional model and its limits in chapter four.

In addition to contemporary dualism, new wave materialism, and reductive functionalism, there are emergent theories of mentality. These are the topic of chapter five. Emergentists agree with functionalists in holding that mental properties and physical properties are distinct. But they disagree with functionalists because they deny that the causal power of mental properties can be reductively explained. Emergentists believe that mental properties have causal power of their own, power that is not simply the result of having physical realizations. In chapter five, I will give my reasons for thinking that emergentism is closer to the type of theory we need for an acceptable solution to the problem of mental causation and for an account of the relationship between mind and body.

CHAPTER 2

SUBSTANCE DUALISM

2.1 Substance Dualism and the Pairing Problem

The mind-body debate generally encompasses two concerns. The first concern is about the metaphysics of conscious beings, whether the ordinary facts associated with a person's mentality requires only one component, such as body, or two components, such as mind and body. The second concern is, supposing a person's mental features are in some way distinct from bodily features, how do they interact? The first concern gives rise to the debates between physicalism and dualism, whereas the second concern represents the problem of mind-body interaction. My dissertation will focus on the second concern.

Dualism at first was seen as substance dualism. Generally, it holds that a body is a physical entity and mind is a mental entity irreducible to any part of a body. When a person experiences pain, for example, his body may incur certain physical changes and these changes cause him to consciously experience pain. When I say that a person is experiencing pain, I mean that a certain qualitative phenomenon is occurring in his mind. The physical changes in the body and the phenomenal changes in the mind are distinct events. The mind and body are distinct substances, yet together they constitute a single person. For substance dualists, mind and body closely interact with each other although they are intrinsically different substances. Mental causation is either the interaction between two mental states within the same mind or the interaction between a mental state and a physical state within the same person. Between these two kinds of interaction, the second one draws more attention.

When we think of substance dualism, we often think first of Descartes. He claimed that mind is essentially different from body although a person is a composite being of the substance of mind and the substance of body. For example, you are aware of that you are thirsty, and you want something to drink. You remember that you bought a Coke yesterday. Then you go to the kitchen and get that Coke. In this account, there are four mental states: you feel thirsty, you want something to drink, you remember that there is a Coke in the refrigerator, and you believe drinking Coke can ease your thirst. Together they bring about your action. Feeling, wanting, remembering, and believing are four different mental states and they occurred only in your mind so they do not have extension in space. Assuming that substance dualism is right and bodies can be changed by changes in minds, how could a non-extended mind come to cause a bodily movement? This is a reasonable question to ask a substance dualist. Descartes' answer depended on the pineal gland, because he believed that it is the locus of mind-body interaction. But we can continue to ask: how could the pineal gland, a physical object, mediate between a mind and a body? A mind is non-material and non-extended so it does not have a spatial location, whereas a body is an extended object. If a mind controls a body through the pineal gland, and if all causation is mechanical in nature, there must be contact between the mind and the pineal gland for the pineal gland to control the body. How could an immaterial mind which has no spatial location contact the pineal gland which has a specific location within the physical world? Descartes himself, it is commonly alleged, never seemed able to satisfactorily address this issue, and the problem of mind-body interaction has continued ever since.

Recently, Jaegwon Kim has offered an explicit argument for the conclusion that such mind-body interaction is incoherent. He calls this argument "the pairing problem." He formulates this problem as an argument to claim that the causal connection between immaterial minds and extended bodies is unintelligible. We are unable to discover any appropriate method to pair a particular mind with a particular body. Given realism about mental causation, the pairing problem is a reason to reject substance dualism. To make sense of this particular problem, Kim begins by looking at the general notion of a pairing relation between a cause and an effect. He claims that a pairing relation underlies causal relationships. If two candidates appear equally good for being a cause of a certain effect, a pairing relation can identify which of the two is the cause of the effect. We generally can think of two methods that could provide a pairing between a cause and an effect. The first method is to trace a continuous causal chain between two events. The second method is to identify a relation that holds between causes and effects. These two methods are related. Let us begin with Kim's simple example of physical causation¹¹: two guns, A and B, are simultaneously fired, and this results in the simultaneous death of two people, Adam and Bob. What makes it the case that the firing of gun A caused Adam's death and the firing of gun B caused Bob's death? Following the first method, we can trace a continuous causal chain between the firing of gun A and Adam's death, and another causal chain from the firing of gun B to Bob's death. But at each link in these chains the pairing problem emerges again: by virtue of what is an event in the A chain paired with the preceding event in the A chain rather than the corresponding event in the B chain, and

¹¹ Jaegwon Kim, *Physicalism*, or *Something Near Enough*. (Princeton University Press, 2005), 79.

by virtue of what is an event in the B chain paired with the preceding event in the B chain rather than the corresponding event in the A chain? In order for a causal chain to explain the connection between the firing of gun A and Adam's death, the pairing problem must be solved. Therefore, Kim dismisses the first method because he thinks that the idea of a causal chain itself presupposes a pairing relation between causes and effects. Assuming the existence of a causal chain to solve the pairing problem begs the question. He says,

... the two methods above,..., are not independent, and this for a very simple reason: the very idea of a causal chain makes sense only if an appropriate notion of causation is already in hand, and this requires a prior solution to the pairing problem.¹²

No causal chain exists from the firing of gun A to Adam's death until we can pair the firing of gun A with the path of the bullet that led to Adam's death, not with that path of the bullet that led to Bob's death. The first method requires the second method. Thus, Kim concludes that only the second method can give us a satisfactory understanding of causes and effects. There must be some relation between two events so that we can pick up the correct pairing relation. If gun A causes Adam's death because there is a pairing relation involved, then he says,

We look for a "pairing relation", R, that holds between A's firing and Adam's death and between B's firing and Bob's death, but not between A's firing and Bob's death or B's firing and Adam's death. In this particular case, when the two guns were fired, gun A, not gun B, was located at an appropriate distance from Adam and pointed in his direction, and similarly with gun B and Bob. It is these *spatial relations* (distance, orientation, etc.) that help pair the firing of A with Adam's death and the firing of B with Bob's death.¹³

¹² Jaegwon Kim, *Physicalism, or Something Near Enough*. (Princeton University Press, 2005), 83.

¹³ Jaegwon Kim, *Physicalism*, or *Something Near Enough*. (Princeton University Press, 2005), 79.

Kim suggests in this situation, that the pairing relation between cause and effect is established by virtue of spatial relations (distance, orientation, etc.) Gun A was oriented toward Adam and not Bob, and it was located at an appropriate distance from Adam whereas gun B was not. Thus, for Kim, the pairing problem between events can be solved when the events have appropriate spatial characteristics.

Now, we turn our attention back to substance dualism. Imagine that there are immaterial minds and the minds have the power to cause bodily movements. If substance dualism is true, there must be a pairing relation underlying the causal relationship between a mind and a bodily movement. Can spatial relations establish the pairing needed for mind-body interactions? Kim says no.

What relation might serve to pair soul A's action with the change in M, a relation that is absent in the case of soul B's action and the change in M? That is, what could be the pairing relation in this case? Evidently, no spatial relations can be invoked to answer this question, for souls are not in space and are not able to bear spatial relations to material things. Soul A cannot be any "nearer" to material object M, or more propitiously "orientated" in relation to it, than soul B is. 14

Minds are not in space and do not bear any spatial relationship to anything. Soul A cannot be any "nearer" to material object M. Soul A and material object M are two different kinds of substance, a mental substance and a physical substance. Solving the pairing problem for mind-body interactions requires that the pairing relation must be some kind of psychophysical relation. But what could it be? It is difficult to imagine what kind of psychophysical relation might be inserted between a mental substance and a physical substance. The whole direction turns back to the original point: the difficulty of

_

¹⁴ Jaegwon Kim, *Physicalism*, or *Something Near Enough*. (Princeton University Press, 2005), 80.

Cartesian dualism. For his part, Kim claims that he cannot think of a way the pairing relation is possible. If a pairing relation is not possible, then it is not reasonable to believe that there is a causal interaction between a mental substance and a physical substance.

And if it is not reasonable to believe that there is a causal interaction between a mental substance and a physical substance, then substance dualism is not reasonable to accept.

Some contemporary substance dualists claim that the pairing problem is not an insurmountable one for substance dualism. They challenge a certain assumption of Kim's argument. John Foster and Peter Unger think that minds and bodies could have dispositions to interact in special ways within particular individuals. W. D. Hart assumes that a mind falls within the same spatial coordinate system as a body. I examine their proposals in the sections that follow.

2.2 John Foster

John Foster, one defender of contemporary substance dualism, argues that popular accounts of physicalism cannot explain the subjective character of our experience.

Assuming that physicalism is right, we are to understand mental phenomena through one of two reductionist strategies. Either we reduce mental concepts to physical concepts or we identify mental states with physical states. The representative conceptual reduction is analytic functionalism, which I will discuss in chapter 4, and the representative of physical reduction is the identity theory, which I will discuss in chapter 3. Foster's strategy is to defend substance dualism by arguing that both reductionisms have to assume the legitimacy of mental phenomena. Analytic functionalism claims that propositions about mentality can be paraphrased into propositions about functional states.

However, according to Foster, the paraphrases fail to capture the subjective character of experience.¹⁵ Foster uses Frank Jackson's famous argument, the knowledge argument, to defend his claim that qualitative phenomena cannot be functionalized. He argues,

In this line of argument, we focus on the case of someone who, owing to a systematic deficiency in his psychological repertoire, does not, for a certain category of mental states, have any introspective data from which he can derive a knowledge of their experiential character. We then claim that, contrary to the implications of analytical functionalism, the subject cannot acquire the relevant experiential knowledge from information about the functional roles of these states in the causal systems of those who have them. ¹⁶

Based on this argument, someone who is born blind has no introspective knowledge of what it is subjectively like to see a red color. Even though he could become a world expert on the physics of color, such knowledge does not help him to acquire introspective knowledge of what it is like to see. So, functional descriptions leave out the subjective features of experience.

On the other hand, the identity theory claims that every mental fact can be identified with a physical fact. Foster does not think that this metaphysical approach works either. To agree with the idea of metaphysical reduction, we have to agree that a color is merely constituted by physical molecules. It follows that the observer has only a brain activity. There is no mental activity at all. Instead of retaining our common-sense notions of mentality, the identity theory forces us to deny mentality altogether. It implies that fundamentally physical objects never have colors for instance. Our experience is in fact illusory. Foster thinks this is a ridiculous consequence because we cannot eliminate

¹⁶ John Foster, *The Immaterial Self*. (Routledge, 1991), 63.

¹⁵ John Foster, *The Immaterial Self.* (Routledge, 1991), 62.

mentality from our experiential viewpoint. I cannot just deny that I am having a visual experience of a table being brown. I cannot just say that I am not feeling a pain in my foot. When I say, "I am happy," there is something more than a mere brain activity. Therefore, no one can deny his own mentality. Foster says,

Whatever conclusions we may reach about the content of the metaphysically fundamental reality, we could never interpret them as explaining *away* the evidence (if 'evidence' is the right term) on which our ascriptions of mentality to ourselves are founded, and thus as representing those ascriptions as unwarranted. In this respect, our ordinary acceptance of our own mentality is entrenched in our system of beliefs in a way that our acceptance of physical-color facts, however natural, is not.¹⁷

Because both conceptual reduction and metaphysical reduction fail, we have to accept the dualist conception of mentality. Foster also argues that we have to accept substance dualism rather than property dualism. If something is just an ordinary physical object, whose essential nature is purely physical, there seems to be no way of understanding how it could be the basic subject of the mental. Foster says,

The problem for this claim, as I have said, is the difficulty of seeing how two such different natures could be co-instantiated – how something could both be a physical object and have, irreducibly, intrinsic properties extraneous to its physical nature.¹⁸

If something is merely a physical object, any understanding of how it is equipped to be a mental subject will have to be explained by understanding its physical nature. But focusing on an object's physical nature will only reveal how it is to engage in activities as a physical object. There is no clue how a physical object can be a basic subject of mental activities. On the other hand, if one proposes that a physical object is equipped to possess

¹⁷ John Foster, *The Immaterial Self.* (Routledge, 1991), 151.

¹⁸ John Foster, *The Immaterial Self.* (Routledge, 1991), 209.

mental features, the difficulty is in understanding how such an object could be. Thus, it has to follow that a basic subject is thought of as having two sides of their natures. It is a corporeal object which also includes a mental component. The mental component is not an aspect of its physical character, but they are equally essential to the basic subject.

If the failure of physicalism shows that physical subjects cannot have any mental attribute, then the subjects are not wholly physical. This corresponds to our ordinary sense of "person." When we say that Jones is in pain, Foster says that generally we are inclined to accept three propositions.

First, the pain belongs to a basic subject; in other words, whatever the possibilities for conceptual and metaphysical reduction, there being something which is *in* pain is an irreducible feature of the situation and requires recognition in the philosophically fundamental account. Second, the basic subject who suffers the pain is Jones himself; that is, the fundamental psychological description of the situation is simply that Jones is in pain at the relevant time. Third, as a *human* individual, Jones has a corporeal nature: he is not a purely spiritual entity, like an angel or a disembodied soul; he is, whatever else, a member of an animal species (*homo sapiens*), with shape, size, and material composition; a solid occupant of physical space; ...¹⁹

Jones is a subject "possessing" something irreducible and at the same time Jones is a subject occupying a certain space and time. For this reason, we may find ourselves attracted to a position that regards subjects as non-physical and also as having corporeal natures. In order to be consistent with a dualist conception, the basic subjects are thought of as having two sides to their natures. On the one hand, they are corporeal objects with shape, size, and other material compositions. On the other hand, their natures also include a psychological component. Thus, Foster makes such a clear dualist claim:

¹⁹ John Foster, *The Immaterial Self*. (Routledge, 1991), 207.

Perhaps we should say that our ordinary concept [of a person] represents a person as something which *both* qualifies as a basic mental subject *and* possesses a corporeal nature, ...²⁰

A person is a dual-natured entity: a basic mental subject, or soul, and a body. These two natures are attached in a special way that allows them to causally affect one another. How can we understand their attachment in a way that makes their causal interaction intelligible?

There is an assumption which Foster says that we use for explaining causal relations. He calls this assumption the "nomological assumption." The "nomological assumption" says that a causal relationship between two events is constituted by lawful relationships between their non-causal properties. For example, we know that heating a lump of metal causes it to melt. What lies behind this causal relationship is a law of nature: whenever a metal of some type reaches a certain temperature, the metal melts.

If the nomological assumption is right, it should explain every case involving a causal relation. However, Foster presents an example to challenge the nomological assumption. Let us say that there is a law of nature, when a lump of metal K reaches a specified temperature, a flash will occur in region R. Based on this law, the causal relationship between metal K's temperature and the occurrence of a flash in R is determined. Now there are two lumps of metal K whose regions of causal influence overlap. Suppose both lumps of metal reach the specified temperature simultaneously and then two flashes occur within the overlapped region. Foster claims that the law does not tell us which lump of metal causes which flash. Thus, "...we have a situation in which

²⁰ John Foster, *The Immaterial Self.* (Routledge, 1991), 239.

the causal relationships transcend the factors which would wholly constitute them if the nomological assumption is true."²¹ Foster thinks something more is needed to connect each lump of metal to a particular flash. Foster's example is an instance of the sort of pairing problem we encountered above. According to Foster, we should not be surprised to discover that causal connections between mental and physical events cannot be explained by the nomological relations between non-causal properties. As his example illustrates such relations are not sufficient to explain the pairings in an ordinary case of physical causation. The nomological assumption alone does not solve the pairing problem.

In the case of causal connections between mind and brain, Foster appeals to a notion of embodiment to supply what the nomological assumption does not. He says,

For the particular brain and the particular mind are linked by some special (and presumably relatively permanent) psychophysical arrangement, which ensures that *this* brain only directly interacts with *this* mind and that *this* mind only directly interacts with *this* brain. Indeed, such an arrangement, as we noted, is required for embodiment: it is, or is an essential part of, what makes it the case that the brain and the mind belong to the same person.²²

To say that a subject is embodied means that a mental subject has a body that is uniquely hers. It is a mistake to look for some third thing to pair a mind and a body since there is a natural bond between the two. Foster claims that a "person" in an ordinary sense is a subject with a corporeal nature. For Foster, this link between a person and her biological organism is fundamental and might exist even before there is a body to interact with,

...it is clear that the connection between the non-physical subject and the biological organism is, fundamentally, a functional one: it is a matter of there being a psychophysical arrangement whereby each partner is

²¹ John Foster, *The Immaterial Self.* (Routledge, 1991), 171.

²² John Foster, *The Immaterial Self.* (Routledge, 1991), 171.

equipped to have the right sorts of direct causal influence on the other. ...It should be noted that, although it equips the subject and the organism to interact, the arrangement may pre-date the time when such interaction is physically possible; for the relevant psychophysical laws may be in place before the organism's neural structures have become sufficiently complex to feature in psychophysical causation. One plausible view, indeed, would be that the arrangement gets established at biological conception, when the organism comes into existence.²³

Foster thinks that this direct functional connection helps a person to recognize that she herself is not only a mental subject but also equipped for interactions between her mental activities and her biological organism. That is to say, when I am aware of "I," I already recognize myself being not only a mental subject but also a body I control. Therefore, the concept of 'I' automatically and fundamentally develops the right sort of direct causal relationship between my mind and my body. Foster also believes that the arrangement of this causal relationship of a particular body and a particular mind may "pre-date the time when such interaction is physically possible."²⁴ When I come to exist, the possible connection between my mental states and my bodily movements also come to exist. That is because my personhood requires an immediate connection between my mind and my body since I have to exist at a location and start to experience. When I say that I am in pain, I am a mental subject with both feelings of pain and neural changes inside my body. Because the concept of "person" requires us to pair a mental subject with a body, Foster concludes that the pairing problem is not a threat for substance dualism. We do not need anything more basic to explain this relationship.

²³ John Foster, *The Immaterial Self.* (Routledge, 1991), 262.

²⁴ John Foster, *The Immaterial Self.* (Routledge, 1991), 262.

Foster, in fact, does not think that the pairing problem exists. The pairing relationship is assumed in the notion of embodiment. However, the problem for Foster's theory is that given the notion of embodiment, it is not at all clear how he facilitates mind-body interaction. The notion of embodiment yields an immediate pairing relation which is unacceptable. Foster seems to claim that the fundamental arrangement of pairing is exactly the embodiment. That is what he needs for the pairing relationship, but he does not provide any details how embodiment pairs. According to him, when I came into existence, my mind and my body just automatically and mysteriously clicked together. But he does not explain further why this particular mind is my mind and this particular body is my body. It seems that there is something needed to explain why a particular mind is embodied in a particular body, and if there is anything more basic needed, the notion of embodiment is not the most fundamental arrangement. That "something" is what we need for the answer of the pairing problem. Unfortunately, Foster just claims that is how embodiment works and we cannot ask for more explanation because that is the most fundamental feature of "person." We still have no understanding from his theory why a particular mind causally connects to a particular body.

2.3 Peter Unger

Peter Unger's argument for substance dualism begins with the 'Experiential Problem of the Many.' It begins by noting that in this world, there are many experiencing thinkers doing different things at the very same time. Each thinker has his own conscious mental life running in parallel simultaneously. It is clear that I know right now that you are not experiencing in a way precisely like the way I am experiencing. I am the

experiencer of my stream of experience. Suppose that substance dualism holds. I as one experiencer, causally interact with one and the same body, *my* body. While my body is typing on a laptop, there are different bodies doing various things. All of them are candidates to be *my* body. Unger claims that there is a relationship connecting my mind with my body. Furthermore, if this world is composed of innumerable molecules, a body is actually nothing more than a collection of molecules that changes continuously. Thus, what we call "my body" is actually a collection of many overlapping bodies. No single member of this collection has any stronger claim for being my body than any other member of the collection. Consequently, in my situation now, there is a mind (the experiencer) interacting with each of many overlapping bodies, each of *my* bodies. Yet, according to Unger, all the bodies in my collection serve to promote *one* experiencing thinker, me. He says,

All the bodies in my situation, however numerous and overlapping, will serve to promote only me myself, and not any other sentient self. And, it may be that just this one causally promoted self is the only experiencer directly interacting with any of these promoting bodies. So, on such a Substance Dualism, there's that motivation to treat *each* of very many bodies as being *one of my bodies*.²⁵

According to Unger, I have many overlapping bodies and none of them is someone else's body. Each of those bodies "causally *promotes just the same single experiencer, just a certain single being that's engaged in just a single (present process of total) experiencing, namely, my total experiencing.* Of course, this experiencer is nobody, but me." Each 'person' is promoted from her own bodies, and each 'person' serves her own stream of

²⁵ Peter Unger, All the Power in the World. (Oxford University Press, 2006), 379.

²⁶ Peter Unger, All the Power in the World. (Oxford University Press, 2006), 380.

experience. Unger believes that this substance dualism would provide a singular resolution for the 'Experiential Problem of the Many' in the sense of a single experiencer being promoted from numerous and overlapping bodies. He also argues that reductive physicalism is not able to explain how one single experience has many overlapping bodies. The doctrines of reductive physicalism reduce one mental property to one physical property. It is difficult for advocates of reductive physicalism to determine which body is the reducing body. Thus, we should be in favor of substance dualism.

According to Unger, although minds and bodies are separately substantial things, one mind is *causally promoted* by an appropriate configuration of many bodies. 'Promotion' is Unger's solution to the pairing problem, and the concept of 'direct propensities' determines which mind is promoted. In other words, the concept of 'direct propensities' determines personhood and the concept of 'promotion' explains the interaction between an experiencer and her bodies. Unger says,

You ask, "How can it be that my body interacts with just my immaterial mind, and not also with (or even only with) the always very similar soul of My Duplicate, always so similar in all Generalistic respects, whose Generally (or Generalistically) so similar body is many galaxies removed from my body?" (It can't be that my immaterial mind is close to my body, of course, since this nonspatial soul hasn't any spatial location at all.) Well, the answer lies in our idea of Individualistically-directed Propensities. Among all the World's many immaterial minds, your body is Propensitied with respect to only the one mind that's your mind, and not with respect to your Duplicate's mind; whereas your Duplicate's body is Propensitied with respect to only his mind, and not with regard to yours. And, for its part, of course, your mind is Propensitied with respect to just your body, not any body but yours, while your Duplicate's mind is Propensitied for interaction with just his body. That being so, this will also be so: When your body is stuck with a pin, it's just with you, or it's just with your immaterial mind, that there's a manifestation of an Individualisticallydirected Propensity to feel pain, whereas (suppose) your currently pin-free Duplicate will be, at this time, pain-free, as well. Now, when you want to assuage this pain, this (Propensity) desire of yours will serve to influence just your own body, and not your duplicate's body. For, this other

Propensity of yours also will be, of course, aptly Individualistically-directed. ²⁷

Suppose that direct propensities determine that a person becomes such person, and a person's mind is causally promoted by particular overlapping bodies based on his own direct propensity. Further suppose we have two people in the same room. In that room, there are two minds, Mary and Jane, and two collections of bodies, Mary-b and Jane-b. Mary-b is sitting on a chair and typing while Jane-b is also sitting on a chair and writing. When Mary-b is doing a physical function, the configuration of Mary-b promotes Mary not Jane. What lies behind the promotion, according to Unger, are Mary's direct propensities. The idea of Individually-directed Propensities solves the pairing problem.

A problem is that Unger simply assumes the existence of direct propensities.

Mary is directly propensitied with Mary-b because when Mary-b is typing, it is just Mary who experiences her body typing. But that appears to be no base which to distinguish Mary's direct propensities from Jane's direct propensities. What is it about Mary's direct propensities that establish the connection between Mary-b rather than Jane-b? Unger does not explain how direct propensities work and this leave the pairing relation unexplained. Furthermore, Unger's solution is essentially similar to Foster's. Both beg the question of how pairing is accomplished by presupposing some type of a special relationship—embodiment for Foster's, direct propensities for Unger—that connects a single experiencer to her body and to no others. When a person comes to exist, the required relation comes to exist and the pairing is finished. Unger, like Foster, never explains how the special relation accomplishes this amazing task.

²⁷ Peter Unger, *All the Power in the World*. (Oxford University Press, 2006), 246-247.

2.4 W. D. Hart

The last contemporary substance dualist I will discuss is W. D. Hart. Following Descartes's strategy, Hart's idea is based on two Cartesian claims:

- (1) If you can imagine that P, then it is possible that P.
- (2) You can imagine that you should be disembodied.²⁸

In defense of (1), Hart argues that imagination is the basic faculty of "how we can know what is possible." Imagination is the faculty we use to know what is possible. If we can imagine that we can walk on our hands, it is possible that we can walk on our hands. If we can imagine that we are disembodied, and of course we can imagine that, then it is possible that we are disembodied. From these two claims, we can conclude that it is possible that we are disembodied. Hart thinks that this conclusion requires a version of dualism. It implies this proposition: anyone who could be disembodied is not purely physical. This proposition, for Hart, is the key insight supporting substance dualism.

Our bodies are located in space, but since a disembodied person has no body, where is he? Hart thinks that any person, even a disembodied person, who exists in this world, must be somewhere. The senses can help us to provide an answer. Every experience of a scene must be from a point of view. Let me take vision in particular. When you are standing in front of this desk and having a visual experience of this desk, you see this desk from the front and the top. You see the flat top, not the bottom. You cannot see the drawers because you are standing on the opposite side. All these visual experiences come with a point of view and where you are located is relative to how you

26

²⁸ W. D. Hart, *The Engines of the Soul*. (Cambridge University Press, 1988), 5.

see this desk. The same think holds true for other types of sensory experience. It seems that all phenomena of sensible experiences require a location in the physical world. Hart thus argues that a disembodied self must have a location within a body. He claims that although I am a disembodied self, I perceive the material world from a particular point of view, *my* point of view. A mind (a disembodied person) should be located in this material world at the particular point of view from which the world is experienced and this demands that the mind is literally located within the body. For Hart, the fact that a mind and a body have the same spatial location provides the key to solving the pairing problem.

Hart believes that causation is a flow of energy, and thus he believes that energy flow is able to explain mental causation. Determining the flow of energy is the method we use to pair a cause with its effect. The energy of the cause decreases and the energy of the effect increases. When energy is traceable, it identifies a causal chain. For centuries, scientists tried to use traceable qualities of energy to identify causal chains. Hart believes that the flow of energy is able to extend to different kinds of entities, even between mind and body. Although mind and body are different kinds of entity and operate using different kinds of energy, the principle of conservation of energy still holds. Whenever some quantity of psychological energy "converts" into a quantity of physical energy, the transformation is subject to the conservation of all energy.

Each person is a disembodied self, so we can locate each disembodied person at his or her particular point of view. Now, suppose that you see a glass of water on a table and a visual experience occurs. Your visual experience involves convergence of lines of sight and those lines spread out from you. When light arrives from that glass along your lines of sight, your visual system loses some physical energy which is converted into

psychological energy. That psychological energy produces the subjective experience of seeing water as well as a belief that it is water that you see. When this occurs, some of the energy from the experience and belief produces a desire to drink the water. When the visual experience continues, the quantity of desire increases. When the energy of desire becomes great enough, it causes you to raise your arm to reach for the water and drink it. Then, the quantity of desires decreases, and the physical energy of a bodily movement increases. The principle of conservation of energy still holds when psychological energy converts into physical energy. Hart thinks that our mental life is full of psycho-physical energy flow between our disembodied selves and bodies.

Although Hart constructs an argument for substance dualism, a pairing between a disembodied self and body is never a problem for him. Each of us can imagine being a disembodied consciousness, and we are non-physical entities lodged in our bodies. That is, our embodied selves are literally located within our bodies, and each embodied self perceives from a particular point of view. The fact that minds and bodies share a location secures the pairing relationship, and the function of conservation of psycho-physical energy governs mental-physical causation.

Given his Cartesian intuitions, Hart argues that if we can image that we are disembodied, it is possible that we are disembodied. He does not provide any further materials to make actuality from possibility. But let me agree with him for a moment: a disembodied self and her body are two distinct substances with the same location. But does having the same location pair a mind to a particular body, or does the existence of a mind-body pairing fix the location of a particular mind? If the first option is right and the pairing problem is solved by locating a mind within a body, how does Hart assign a

location to a mind since space is not included in a mind's essence? He cannot simply assume that mind is located in the body that causes the visual experience, because the location of the mind is supposed to determine the body with which it causally interacts. If Hart uses the existence of a mind-body pairing to fix the location of a mind, which is the second option, there must be something more basic that explains the pairing of a mind and a particular body. However, Hart never provides that more basic principle to explain the mind body pairing. Thus, the pairing problem remains. Either way, Hart does not make clear how the pairing relationship is made.

2.5 Conclusion: the Problem of Mental Causation Unsolved

The problem of mind-body interaction has plagued substance dualism ever since Descartes. The pairing problem is simply the most recent articulation of that problem. Recall that Kim argues if there is a causal relation between mental and physical events, then there must be a way for us to pick up the right pairing relation instantiated by the causal relation. If minds and bodies are independent substances and involved in a causal relation, there *must* be a pairing relation associated with them. Discovering this pairing relation is a critical task for substance dualists. If the only thing that can play the role of establishing the correct pairing relation is the spatial relation, then it seems to follow that the possibility of the causal relation between minds and bodies in substance dualism is precluded. Thus, Kim concludes that the only kind of relation that explains the pairing relation cannot be intelligibly held by substance dualism. The problem of mental causation is unintelligible in substance dualism.

In response to Kim, John Foster thinks that the relation of embodiment explains that a person is a basic mental subject and also has a corporeal part. Given embodiment, the pairing problem is not a problem. The pairing relationship is already established when a person, a single subject with both a mental and corporeal nature, comes into existence. When a person comes to be, she has both a body and a mind. In other words, when a person comes to be, her mind is already embodied. 'Embodiment' provides an immediate pairing between a particular mind and a particular body. I have argued that it is unacceptable to use embodiment to assume the functional arrangement of a mind and a body. Foster never tells us how embodiment works and he avoids this question by claiming that the arrangement predates the actual existence of a body. However, why this mind is *my* mind and this body is *my* body is still mysterious. Something more basic is needed to explain embodiment. Something Foster does not provide.

Peter Unger thinks each mind is promoted by the appropriate configuration of billions of bodies. In other words, billions of bodies working together promote a single experiencer. So *my* mind is paired with *my* body because I am aware that it is me doing what I am doing while the whole bodily system is doing the same thing and further, I know no one else is doing precisely what I am doing. By saying this, Unger wants to claim something intrinsic to every individual. The intrinsic qualities to each mind are "direct propensities." The direct propensities determine I am the individual possessing my own experience stream, and it involves reciprocal parts, a mind and a body. In Unger's account, he uses Individualistically-directed Propensities to pair one particular mind and one particular body because those propensities necessarily require the interaction between minds and bodies. If Unger's theory is true, a particular mind is

paired with billions of bodies by direct propensities. Direct propensities determine who I am, but he does not explain why this particular mind is *my* mind and that group of bodies is *my* body. I have argued that Unger's solution to the pairing problem is similar to Foster's solution. Both Foster's notion of embodiment and Unger's notion of direct propensities presume a special relationship between a mind and a particular body at the moment a person comes into existence. However, neither the notion of embodiment nor the notion of direct propensities answers the pairing problem. Both philosophers try to "solve" the pairing problem by assuming that when a person comes to exist, *her* mind and *her* body will be attached automatically.

W. D. Hart's dualism assigns spatial location to a particular mind, because each mind perceives the world from a unique perspective or point of view. Thus, a mind must be located in the physical world at the same location of its body. Since it is one point of view, Hart assumes that a mind must be located within a body, and it pairs a mind with a body. If minds like bodies are located in space, the shared location makes possible a psycho-physical energy flow. If Hart's theory is true, we have to assume a mind's and a body's location before a person starts to perceive the world. We have to assume a particular mind was already within a particular body before that person experiences the world from any particular point of view. This means that location makes mental-physical causation happen. However, location is not a mind's essential property, so Hart needs another story to explain how a mind is located within a body prior to any mental-physical interaction.

Like Descartes, contemporary substance dualists face the pairing problem, which makes mind-body interaction unintelligible. I conclude that we do not have any reason to

accept substance dualism because no version of it has solved the problem of mental causation.

CHAPTER 3

THE MIND-BODY IDENTITY THEORY

As we have seen, substance dualism leaves both the pairing problem and the problem of mental causation unsolved. Some philosophers have tried to respond to these problems by suggesting that we give up the existence of mind as a substance and locus for mental facts. Persons have minds, but what is it to have a mind? We perceive, we sense, we even believe and desire. We have all these mental features, but what is their nature? The mind-body identity theory shows that there is no need to multiply entities beyond necessity. Mental properties are identical with brain properties. If the identity theory is true, mental causation is nothing over and above physical causation so the problem of mental causation will have been solved. But in recent years, there have been strong philosophical arguments raised against the possibility that the mind-brain identity theory is true. In section 2, I will consider Kripke's philosophical argument against it. In section 3, I will explain contemporary replies to Kripke, replies that have made the mind-brain identity theory a serious contender in the philosophy of mind. Finally, in section4, I will consider whether there are good arguments in favor of the identity theory.

3.1 The Identity Theory --- The Old School

The possibilities of mind-brain identity began to be accepted as a serious philosophical thesis after several papers published in the late 1950's. The most influential

were papers by U. T. Place and J. J. C. Smart.²⁹ They argue that mental events and mental processes are not merely correlated with events and processes in the brain: they are the very same events and processes as those of brain activities. Their theory became known as the "mind-body identity theory."

Although Place and Smart discussed different aspects of our mentality, consciousness for Place and sensation for Smart, they were both defending the same basic position. In "identifying" mental states with brain states, Place spoke of the identity of constitution rather than of an identity of meaning. One of his typical examples is "A cloud is a mass of water droplets." The meaning of "a cloud" is obviously not the same as the meaning of "a mass of water droplets." The experience by which I see a white object up in the sky and the knowledge by which I understand that a cloud is a mass of water droplets are different. The former is a phenomenal experience and the latter is a scientific discovery. However, the difference between these kinds of knowledge does not imply that their referents are two different objects. Place said,

It is clear from this that the terms 'cloud' and 'a mass of tiny particles in suspension' mean quite different things. Yet we do not conclude from this that there must be two things, the mass of particles in suspension and the cloud. The reason for this, I suggest, is that although the characteristics of being a cloud and being a mass of tiny particles in suspension are invariably associated we never make the observations necessary to verify the statement 'that is a cloud' and those necessary to verify the statement 'this is a mass of tiny particles in suspension' at one and the same time. We can observe the micro-structure of a cloud only when we are enveloped by it, a condition which effectively prevents us from observing those characteristics which from a distance lead us to describe it as a cloud.

²⁹ See U. T. Place, "Is Consciousness a Brain Process?" *British Journal of Psychology*, Vol. 47, (1956) and J. J. C. Smart, "Sensations and Brain Processes." *The Philosophical Review*, Vol. 68, No. 2, (1959).

Indeed, so disparate are these two experiences that we use different words to describe them.³⁰

According to Place, the 'is' in the statement, 'A cloud is a mass of water droplets,' does not function as a definition. 'A mass of water droplets' does not define 'a cloud.' So this statement is not one of analytic necessity. The truth of this statement is verified by scientific discovery. In other words, this statement is true a posteriori. To verify the 'is' between a cloud and a mass of water droplets, the observations of each are required. By scientific technology, we can verify that the expressions 'a cloud' and 'a mass of water droplets is suspension' refer to the same thing although their observation conditions are different. Place thought that the statement 'consciousness is a brain process' is also an a posteriori statement. The observation of consciousness and the observation of a brain process occur in very different contexts. Consciousness refers to our subjective experience and statements about consciousness are verified through introspective reports. For example, I am aware of feeling hungry. But knowledge of brain activity and function comes through neuroscience. Place argued that this difference of verification conditions is consistent with understanding consciousness and brain activity as the same phenomenon.

Place argued that to believe that consciousness and brain processes could not be identical is to be misled in a particular way. When we describe our experience, we believe that our introspective description of it truly reflects the features of some real phenomenal object. Although introspective descriptions do reflect phenomenal properties in our experience, there are no phenomenal objects that have those features. Place pointed

³⁰ Ullin T. Place, "Is Consciousness a Brain Process?" In *British Journal of Psychology*, Vol. 47, (1956): 47.

out that it is fallacious to suppose that there are. Just like the example of a cloud and a mass of water droplets, the observation of consciousness and the observation of brain processes should not mislead us to believe that there are two kinds of entities: phenomenal entities and physical entities. Place called this mistake the "phenomenological fallacy." Place said,

The phenomenological fallacy on which this argument is based depends on the mistaken assumption that because our ability to describe things in our environment depends on our consciousness of them, our descriptions of things are primarily descriptions of our conscious experience and only secondarily, indirectly, and inferentially descriptions of the objects and events in our environments. It is assumed that because we recognize things in our environment by their look, sound, smell, taste, and feel, we begin by describing their phenomenal properties, i.e. the properties of the looks, sounds, smells, tastes, and feels which they produce in us, and infer their real properties from their phenomenal properties.³¹

Place thought that it is wrong for us to believe that the features appearing in our phenomenal field are real and different from properties of brain processes. For example, when a person reports a green after-image, he is asserting what is happening *in* him. In reality, there is no entity in the world which is a green after-image. Nor is there a green object in his brain. A green after-image is a phenomenon that occurs when our eyes receive a specific type of stimulation. When a person reports a green after-image, what he reports is something that is going on in his brain. The experience causes him to think that he is seeing a special sort of object---a green after-image. But there is no such object. The introspective description of a green after-image refers to an aspect of the visual process itself. Any conscious experience is a brain process. We recognize things by their looks, smells, taste, and feel, but it does not mean that we have to assume that those properties

³¹ Ullin T. Place, "Is Consciousness a Brain Process?" *British Journal of Psychology*, Vol. 47, (1956): 49.

are features of a non-physical reality. Therefore, Place believed that mind-brain identity theory helps us to adequately explain a person's observations of consciousness.

J. J. C. Smart followed Place in defending the identity theory, but Smart cited Occam's razor to support his claim. Mental phenomena are different from real objects occupying space and time. Although the mind-brain identity theory and dualism are both consistent with the known facts, according to Occam's razor, the former has an advantage in virtue of its simplicity and explanatory utility. ³² Substance dualism supposes the existence of phenomenal substances that exist independently from bodies. Although dualism does provide metaphysical "space" for conscious experience, it does not help to explain what consciousness really is. Like Place, Smart believes that we posit mental substances because we have introspective descriptions, but such descriptions do not entail the existence of phenomenal objects. According to Smart, when a person says, "I see a green after-image," he means roughly, "There is something going on in me which is like what is going on when I open my eyes and there is something green illuminated in front of me."³³ However, there is no "something" in front of me when I open my eyes. What is occurring in me makes me think that there is something in front of me. The truth is, the occurrence of a mental image is merely a brain process. Research into the brain will eventually explain what sensory experience is, without the need for a separate type of substance. For these reasons, Smart claimed that mind-brain identity theory is a better theory than mind-body dualism.

-

³² J. J. C. Smart, "Sensations and Brain Processes." *The Philosophical Review*, Vol. 68, No. 2, (1959): 142.

³³ J. J. C. Smart, "Sensations and Brain Processes." *The Philosophical Review*, Vol. 68, No. 2, (1959): 149.

Smart argues that the relation of identity between a sensory experience and a brain process is established by the common reference of two kinds of descriptions: descriptions of subjective experiences and the corresponding descriptions of brain processes. Most of us focus on the meaning of mental concepts and brain concepts so much that we do not consider that despite the differences in their meanings, they might have identical referents. Indeed it is possible that a person knows lightning without knowing lightning is an electric discharge. Similarly, when we report a sensation, we need not know that a sensation is a brain process, and we are not reporting a brain process. So Smart said,

"I see lightning" does not *mean* "I see an electric discharge." Indeed, it is logically possible (though highly unlikely) that the electrical discharge account of lightning might one day be given up. Again, "I see the Evening Star" does not *mean* the same as "I see the Morning Star," and yet "the Evening Star and the Morning Star are one and the same thing" is a contingent proposition.³⁴

The meanings we know about "the lightning" and "electric discharge" are different, but different meanings do not establish different referents. Empirical observation is required in order to take that further step. For what is reported in the introspective report is simply an experience of yellow not an entity of yellowness. Thus, the mind-brain identity theory is not to tell us what mental concepts and brain concepts mean but what they refer to.

Two different kinds of concepts both happen to refer to brain processes.

Smart is not claiming that phenomenal after-images and the like are actually brain entities. There is no phenomenal after-image. Phenomenal properties are not real entities

_

³⁴ J. J. C. Smart, "Sensations and Brain Processes." *The Philosophical Review*, Vol. 68, No. 2, (1959): 147.

in the environment or in the brain. It is better to say that the whole process of experiencing *is* a brain process. Smart makes this point clear,

I am not arguing that the after-image is a brain-process, but that the experience of having an after-image is a brain-process. It is the *experience* which is reported in the introspective report. Similarly, if it is objected that the after-image is yellowy-orange but that a surgeon looking into your brain would see nothing yellowy-orange, my reply is that is it the experience of seeing yellowy-orange that is being described, and this experience is not a yellowy-orange something. So to say that a brain-process cannot be yellowy-orange is not to say that a brain-process cannot in fact be the experience of having a yellowy-orange after-image. There is, in a sense, no such thing as an after-image or a sense-datum, though there is such a thing as the experience of having an image, ... ³⁵

Although a phenomenal property may be yellow, that person's *experience* of that property does not *possess* that property. What is reported in an introspective description is simply a report of a process of experiencing a yellow image, not a report of a yellow entity. It is the whole thing of experiencing which is identical with a brain process. If we confuse a report of experiencing a yellow image with a report of a yellow entity, we commit the phenomenological fallacy of which Place spoke. Smart suggested that although ordinary language seems dualistic, the occurrence of an experience does not commit us to dualism. Smart concluded that sensory experience is nothing other than a brain process.

Both Place and Smart advocated the mind-brain identity theory. The claim that consciousness is a brain process is a contingent and testable scientific hypothesis. Smart argued that we can distinguish that something is a flash of lightening or a motion of electric charges in a different way, but this different way does not prevent a flash

39

³⁵ J. J. C. Smart, "Sensations and Brain Processes." *The Philosophical Review*, Vol. 68, No. 2, (1959): 150-1.

lightening being identical to a motion of electric charges. Similarly, 'sensation' and 'brain process' differ in meanings but they have the same referents. The advantage of this theory is that there is no more myth of mind, and there is no such mind-body problem.

3.2 Philosophical Arguments against the Identity Theory

Smart and Place had argued that the mind-brain identity relation could be scientifically verified. Phenomenal properties and scientific properties do not have the same meaning, but that does not imply that the mind-brain identity theory is false. The identity theory tells us that the descriptions of phenomenal properties and the descriptions of scientific properties fix the same referents, and thus the identity relation between mind and brain is neither *a priori* nor necessary. Saul Kripke has argued against this view of mind-brain identity.

Kripke says that a designator is rigid just in case it refers to the same object in all the possible worlds where it has a referent. We say that 'George W. Bush' is a rigid designator, and this name must refer to the same man in all the possible worlds that he exists. Kripke claims that names in ordinary language are rigid designators, but definite descriptions are not rigid. 'George W. Bush' is a rigid designator and always picks out the same object in all possible worlds. It does not matter if it is true that he is the 43^{rd} president of the United States or a father of two daughters in each of these possible worlds. Those descriptions might or might not refer to Bush in a possible world. Yet whatever person these descriptions pick out in different possible worlds, the name 'George W. Bush' always picks out the same person. Generally speaking, if 'x' and 'y' are rigid designators, then if x = y, then necessarily x = y. For example, 'Cicero' refers to

Cicero and 'Tully' refers to Tully, and we know that Cicero and Tully is the same person.

Since Cicero = Tully is true and 'Cicero' and 'Tully' are both rigid designators,

necessarily Cicero = Tully.

Kripke separated the metaphysical concept of necessity from the epistemological concept of aprioricity. Identities can be necessary even if they are not known apriori. Kripke extended his account of rigid designation to include not only proper names but also natural kind terms such as 'water' and 'H₂O'. Scientific statements of identity such as 'water is H_2O ' are also necessary if true. He then asked, why is it that 'water = H_2O ' appears contingent? If 'water' and ' H_2O ' are rigid designators, then necessarily water = H₂O. However, it seems to us that we can easily imagine a possible world in which 'water' does not pick out H₂O. Kripke explains that the cluster of properties commonly associated with water and ordinarily used to identity water, properties such as being transparent, tasteless, and potable, etc. are accidental characteristics of water. A definite description using this cluster of properties which I abbreviate as 'the watery stuff' picks out water in the actual world. But this description, like all definite descriptions is a nonrigid designator. That is, 'the watery stuff' does not necessarily refer to 'water.' Thus, the referent of 'water' is fixed contingently, and there are possible worlds in which 'the watery stuff' does not pick out H_2O , but another chemical compound XYZ. Kripke argues that what is genuinely possible is not that 'water' is not H₂O, but that what 'the watery stuff' picks out is not H_2O . So the apparent contingency of water = H_2O is explained by the fact that we can imagine a possible world where we occupy the same epistemic situation we do in the actual world when surrounded by a large body of the watery stuff. In that situation, it *appears* that we are surrounded by water, yet our

metaphysical situation is in *reality* quite different---we are not surrounded by water, but by XYZ.

But why can we not use the same analysis to explain away the apparent contingency of pain = C-fiber firing? While the referent of 'water' is determined contingently, the referent of 'pain' is determined essentially. The referent of 'pain' is most commonly determined by the phenomenal character of pain --- the feeling of pain. The phenomenal character of pain is essential to it, so the relation between the description, 'the feeling of pain,' and 'pain' is necessary. There is no possible world in which a person is in the epistemic situation described by having 'the feeling of pain' but is not in pain. Thus, there is no situation that parallels being surrounded by 'the watery stuff' but *not* being surrounded by water. In other words, if pain = C-fiber firing is necessary, and if the feeling of pain is an essential property of pain, the description, 'the feeling of pain,' always picks out C-fiber firing.

It is possible that H_2O is not transparent, tasteless, and potable, etc. In other words, the watery stuff does not include any essential features of H_2O . Therefore, the contingency of 'water = H_2O ' can be explained away by the fact that the properties we use to fix the referent of 'water' are ones it possesses contingently, but this is not the case for 'pain.' In the case of water = H_2O , we are misled by the fact that we might have been in an epistemic situation indistinguishable from the one that we typically occupy when in the presence of water without being in the presence of water. But there is no epistemic situation that is indistinguishable from the one that we typically occupy when in the presence of pain that is one in which we are not in pain. In the case of water, the necessary statement 'water is identical with H_2O ' is confused with the contingent

statement 'the stuff that is watery is identical with H₂O.' The possible falsehood of 'the stuff that is watery is identical with H₂O' leads us, mistakenly, to suppose that 'water is identical with H₂O' might also be false. But in the case of pain, both the statements 'pain is identical with C-fiber firing' and 'the feeling of pain is identical with C-fiber firing' are necessarily true if true at all. So, according to Kripke, the only way to explain the contingency commonly associated with 'pain is identical with C-fiber firing' is to conclude, contrary to what the identity theorists claim, that this statement is false.

The mind-brain identity theory claims that mental states and brain states are identical, but the identity relation is contingent. Kripke's argument purports to show that there is no contingent identity. This leaves three possibilities regarding mind-brain identity statements: they are either false or necessarily true, or they are confused with genuinely contingent statements. Since identity theorists such as Smart and Place deny that they are necessarily true, they must be either false or confused with genuinely contingent statements. But Kripke argues that there are no genuinely contingent statements with which they have been confused. Thus, Kripke concludes that mind-brain identity statements are false. However, other fans of the psychophysical identity thesis, calling themselves "new wave materialism," have uncovered new possibilities.

3.3 Christopher Hill, Brian McLaughlin and New Wave Materialism

Thomas Nagel was circumspect about the relation of mental and physical properties in his well-known article, "What is it Like to be a Bat?" He claimed we have various subjective concepts that we acquire by introspection, imagination, sympathy, etc.

These special subjective concepts make mind-brain identity statements fundamentally

mysterious given our current scientific perspective. Yet, Nagel believed that this fact does not entail that mind-brain identity statements are false. In a footnote, Nagel acknowledged the need to explain the apparent contingency associated with mind-brain identity statements. He sketched the following proposal to surmount this problem,

To imagine something perceptually, we put ourselves in a conscious state resembling the state we would be in if we perceived it. To imagine something sympathetically, we put ourselves in a conscious state resembling the thing itself...When we try to imagine a mental state occurring without its associated brain state, we first sympathetically imagine the occurrence of the mental state: that is, we put ourselves in a state that resembles it mentally. At the same time, we attempt to perceptually imagine the non-occurrence of the associated physical state, by putting ourselves into another state unconnected with the first: one resembling that which we would be in if we perceived the non-occurrence of the physical state. Where the imagination of physical features is perceptual and the imagination of mental features is sympathetic, it appears to us that we can imagine any experience occurring without its associated brain state, and vice versa. The relation between them will appear contingent even if it is necessary, because of the independence of the disparate types of imagination.³⁶

Perceptual imagination and sympathetic imagination operate independently of each other. For example, suppose my friend Mary gets a paper cut and is experiencing pain. When I try to imagine how she feels, I try to put myself in a mental state which resembles Mary's feeling. I might do this by recalling as vividly as I can, situation in which I have had paper cuts. This process is an example of sympathetic imagination. By contrast, when I try to imagine how I would look with red hair, I can just look into a mirror and imagine seeing my reflection with red hair. There is no need to imagine myself feeling a certain way. I just imagine myself in a different perceptual state. This is an example of

³⁶ Thomas Nagel, "What Is It Like to Be a Bat?" *The Philosophical Review* Vol. 83, No. 4, (1974): 446, fn. no. 11.

perceptual imagination. According to Nagel, the type of imagination used when we imagine being aware of C-fiber firing is exclusively perceptual; however, the type of imagination used when we imagine being in pain is sympathetic. By distinguishing them, Nagel uncovered another possibility for explaining the apparent contingency of mindbrain identity statements, namely, that we can perceptually imagine that someone is undergoing C-fiber firing without sympathetically imagining that person in pain, and vice versa.

Christopher Hill exploits Nagel's insight in an argument against Kripke. Tirst, Hill and Kripke both agree that according to the mind-brain identity theory, if pain = C-fiber firing, when a person is in pain, C-fiber firing occurs in his brain. Second, both agree that 'the feeling of pain' is an essential property of 'pain.' Now Hill asks us to suppose it is possible to perceive one's own brain activity by using a specialized scientific apparatus. For example, we can imagine that in a laboratory, there is a machine with some cables. When those cables are attached to my brain, I can observe my brain activities through a monitor connected to that machine. Using this machine I could perceive whether or not C-fiber firing were occurring in my brain, Imagining whether or not C-fiber firing is occurring in my brain is like imagining whether or not I have red hair. In both situations my perceptual imagination is at work. Because perceptual imagination is independent from sympathetic imagination, Hill argues that when I perceptually imagine C-fiber firing occurring in my brain, it is not necessary that I am also sympathetically imagining myself in pain. That is, when perceptual imagination occurs, it

³⁷ Christopher Hill, "Imaginability, Conceivability, Possibility and the Mind-Body Problem." *Philosophical Studies*, 87 (1997).

is not necessary that sympathetic imagination has to occur as well. By using this method, Hill thinks that the identity theorists can explain away the apparent contingency of identity statements connecting pain and C-fiber firing.

Hill thinks that there are other imaginative situations that reinforce the apparent contingency associated with mind-brain identities. One person can observe another person's C-fiber firing by watching it through a lab's machine but the observer is not himself in pain. Thus, it is possible that one person can imagine *perceiving* C-fiber firing without actually being in pain. The contingency can be reinforced by the contrary situation. When I am in pain, but not attached to the machine, I am not aware of the brain activity of C-fiber firing. Hill says because we can so easily imagine situations in which we are aware of pain, but not aware of C-fiber firing, and vice versa the apparent contingency of their identity is strengthened. However, that contingency is due to the different imaginative processes at work and not to the metaphysical distinctness of pain and C-fiber firing. There are then two true genuinely contingent statements that are easily confused with the true necessary statement 'pain is identical with C-fiber firing.' Those contingent statements are 'pain is identical with the state which I am perceptually imagining that I occupy' and 'the state which I am sympathetically imagining that I occupy is identical with C-fiber firing.' Thus, the identity between pain and C-fiber firing is a posteriori, as Place and Smart believe, necessary as Kripke maintains, and yet apparently contingent.

Brian McLaughlin has also argued that the apparent contingency of mind-brain identities can be explained without undermining their truth. His approach is slightly different from Hill's. McLaughlin claims that the two terms flanking the identity sign are

in fact different concepts. 'Pain = C-fiber firing' is contingent because 'pain' and 'C-fiber firing' are playing different conceptual roles. When we conceive of a property under a phenomenal concept, we must conceive of this property as a phenomenal property. For example, when we conceive of 'pain,' we must conceive it as a qualitative phenomenon. On the other hand, if we conceived of 'C-fiber firing,' we have to conceive of it as a brain activity. There is no qualitative phenomenon involved. McLaughlin thus claims that the way we conceive 'pain' is different from the way we conceive 'C-fiber firing.' But this conceptual difference is consistent with the claim that pain is identical with C-fiber firing. When we conceive 'pain' and 'C-fiber firing' together, we see no *a priori* reason to believe they refer to the same thing. The act of conceiving is not a reliable guide to identity or non-identity.

McLaughlin argues that failing to keep conceptual roles and properties distinct generates confusion. Pain is a kind of feeling and to conceive it as a feeling is to conceive it differently than as C-fiber firing. A feeling and C-fiber firing differ in their conceptual roles. But there is no reason to believe that this difference in conceptual role implies a difference in the properties referred to. In other words, a conceptual difference is not a reliable guide to a metaphysical difference. Scientific discoveries confirm the strict correlation between phenomenal properties and physical properties, and this correlation is best explained by identifying the two types of properties. Yet this identification is compatible with significant differences in the concepts that refer to this property.

According to Hill and McLaughlin, because proponents of the identity theory can explain the apparent contingency of mind-brain identities by appealing to different types of imaginations, or by separating conceptual roles from properties, new wave materialists

claim that the mind-brain identity theory is not refuted by Kripke's arguments. But to show that a position has not been refuted is not to show that the position is true.

Generally speaking, new wave materialists also argue that the mind-brain identity theory does a better job of explaining mental-physical correlations than Cartesian dualism, epiphenomenalism or a dual- aspect theory. The mind-brain identity theory, they say, offers the best account of our mentality.

3.4 Arguments for Mind-Brain Identities

The identity theory says that for every mental property M, there is a physical property P and these properties are not merely consistently correlated with one another; they are identical. Mental properties *are* physical properties. There is only one property although it is conceived under two different concepts. If it is true that mental properties are identical with physical properties, mental causes are identical with physical causes as well. New Wave Materialists argue that insisting on the distinction of mental properties and physical properties makes causal power of mental properties mysterious. Identity theorists claim there is only one kind of property, physical properties, and that there is "nothing over and above" the physical level. Consequently, it is not surprising that if mental properties are reduced to physical properties, mental causation no longer exists as something over and above physical causation. To say that mental causes are physical causes is to say that there is no distinctively mental causation. All there is is physical causation. The problem of mental causation for the identity theory is simply solved by claiming that mental causation is physical causation.

But the fact that the truth of the identity theory would neatly dispose of the problem of mental causation does not establish that the theory is true. What reasons or arguments have the identity theorists put forward to claim that mind-brain identities are true? There are three types of arguments that have been proposed to provide positive support for the identity theory.

The first is the simplicity argument. Smart argues that mind-brain identity helps us to attain the simplest worldview by appealing to Ockham's razor. One of the general formulations of Ockham's razor says that we should not posit more entities beyond what is necessary. 38 The identity theorists believe that it is not necessary to posit either distinct mental substances or distinct mental properties. They think that a theory which takes mind to be identical with brain is simpler and it can explain the features commonly associated with our mentality. For the identity theorists, mind is not something "over and above" brain. All mental problems are brain problems. In other words, the identity theory is the simplest because the identity theorists believe that there are no private qualitative phenomena such as feelings of pain or yellow after-images. However, dualists think that an appropriate theory of mind should not leave out qualitative phenomena. The identity theory is the simplest because it dismisses traditional aspects of the problem of qualitative phenomena. But that is not what we want for a simplest theory for mental properties. For example, if I ask, what is the simplest way from Carbondale to Chicago? Someone would say that the simplest way is driving. It is simple because there is no waste of time and it is easy to find a car. But another person might say that taking the

_

³⁸ Jaegwon Kim, *Philosophy of Mind*. (Westview Press, 2006), 89.

train is the simplest because you do not need to drive by yourself. What you need to do is to buy a ticket and put yourself on the train. This example shows that the meaning of 'simplicity' is relative to background presuppositions. Some people consider time an issue, so the simplest way to them takes the length of the trip into account. Other people do not want to exhaust themselves and they do not consider the length of the trip a problem, so taking train is the simplest way for them. Similarly, deciding the simplest way to explain mental properties will presuppose background facts. The identity theorists believe that their theory is the simplest because they dismiss the possibility that there are distinctly qualitative phenomena. The dualists do not agree with this background presupposition. Is the *feeling* of pain simply a brain process? We commonly believe that such private qualitative phenomena are not just brain processes. We do not understand how some features of our experience could be physical features of our brains. They are something "over and above" physical properties. While the identity theorists presuppose a physicalist worldview, the dualists think that there is compelling first-person evidence for the existence of non-physical qualitative properties. Simply appealing to the simplicity argument does not help to decide which worldview we should presuppose. Because dualists and identity theorists do not agree on their background presuppositions, Ockham's razor cannot be used to support the identity theory without further independent arguments.

The second positive argument for the identity theory is that it offers the best explanation for mind-body correlation. It is commonly accepted that mind and body are correlated. For example, C-fiber firing occurs every time pain occurs. There apparently is a correlation between 'pain' and 'C-fiber firing.' So every theory of mind needs to

explain this correlation. McLaughlin and Hill both agree that the identity relation is the best way to explain mind-body correlation. Pain is identical with C-fiber firing and it explains the fact that C-fiber firing occurs every time pain occurs. In other words, there is in fact no correlation between mind and body, but only the identity relation. Because they are identical, there is nothing to be explained. The account offered by Hill and McLaughlin, if true, would explain why people have *falsely* believed that mental properties and physical properties are constantly correlated but distinct. But, as we noted above, their accounts presuppose the truth of the identity theory and so cannot be used as positive argument in its favor.

The final argument is that the mind-brain identity theory offers the best explanation of causal chains within folk psychology, such as "pain causes a feeling of distress." Identity theorists claim that there is only one kind of property, physical properties, and that there is "nothing over and above" the physical level. Consequently, it is not surprising that if mental properties are reduced to physical properties, the mental causation no longer exists as something over and above the physical causation. Since everything is fundamentally physical, an event of knee pain causing an event of holding knees is explained at the physical level.

- (1) Neurophysiology
- (2) NP1 causes NP2
- (3) NP1 = Jane's pain
- (4) NP2 = Jane's holding of her knee

(5) Jane's pain causes Jane's holding of her knee.³⁹

According to the identity theorists, neurophysiological theory explains (2) and (2) together with (3) and (4) allow the derivation of (5). Thus, the identity statements in (3) and (4) enable us to explain why Jane's pain causes her to hold her knee. If the identity relation is able to explain why and how a mental property occurs, the causal power of that mental property is able to be explained in terms of the causal powers of physical properties as well. The causal power attributed to the mental state of pain in (5) is transferred by (3) to the causal power of NP1, and the causal power of NP1 is explained through neurophysiological theory.

Jaegwon Kim thinks that the identity relation does not play an explanatory role in this scenario. He says,

An identity reduction of X to Y does not generate an explanation of the occurrence of X in terms of Y, or Y in terms of X; nor does it enable us to explain why X has a certain property F by saying that Y has property F and X = Y.

If the identity theory is true, NP1 = Jane's pain says nothing more than NP1 = NP1.

According to Kim, the identity relation only allows us to *rewrite* "NP1" as "Jane's pain."

Kim argues that such rewrites do not generate any explanatory power of pain. He says,

Identities seem best taken as mere rewrite rules in inferential contexts; they generate no explanatory connections between the explanandum and the phenomena invoked in the explanans; they seem not to have explanatory efficacy of their own.⁴¹

³⁹ I am modeling this example after Kim's on *Physicalism, or Something Near Enough*. (Princeton University Press, 2005), 144.

⁴⁰ Jaegwon Kim, *Physicalism, or Something Near Enough*. (Princeton University Press, 2005), 118.

⁴¹ Jaegwon Kim, *Physicalism, or Something Near Enough*. (Princeton University Press, 2005), 132.

What the identity relation does is merely rewrite the fact that physical relationships have already explained. NP1 has already explained why NP2 occurs, and although identity statements permit us to rewrite that fact as Jane's pain causes Jane's holding of her knee, they add nothing to the explanation. The explanation ends with (2).

Kim grants that NP1 = Jane's pain carries more information than NP1 = NP1 does. At least there is one piece of information in the epistemological dimension. It is the fact that "Jane's pain" and "NP1" refer to the same thing, and Kim thinks that this fact is a proper explanandum. It implies that we can ask why 'Jane's pain = NP1' is true? If true, the statement 'Jane's pain = NP1' is necessarily true, but it is *a posteriori*. In other words, we do not realize that "Jane's pain" and "NP1" refer to the same thing until we discover they are identical. That is, knowing the fact that Jane's pain and NP1 are identical is an additional piece of knowledge different than knowing that NP1 is NP1. But this additional piece of information adds nothing to the explanation of (5). New wave materialists hold that mind-brain identity statements enable us to explain causal relations in folk psychology. But, following Kim, we have seen that such statements only allow us to rewrite causal relations of folk psychology in physical terms. The identity statements themselves contribute nothing to the explanation of the causal relations involved.

_

⁴² Jaegwon Kim, "Reduction and Reductive Explanation: Is One Possible Without the Other?" *Being Reduced: New Essays on Reduction, Explanation, and Causation*. ed. by Jacob Hohwy and Jesper Kallestrup. (Oxford University Press, 2008), 103.

3.5 Arguments against Mind-Brain Identities

By identifying mental properties with physical properties, mind-body theorists hope to show that the mind-body problem is a pseudo-problem. All aspects of mentality can be explained at the physical level, including the phenomena of mental causation. However, Kim points out and I agree that arguments typically offered in support of mind-brain identities are inadequate. Equally important is the fact that serious arguments, other than Kripke's, have been put forward against the identity theory. The two most famous are the multiple realization argument and the qualia argument. We will consider the multiple realization argument here and discuss the qualia argument in one section of chapter four.

The concept of 'multiple realizability' was introduced by Hilary Putnam in the 1960's as a consequence of his functionalist approach to the mind-body problem. Putnam said,

Consider what the brain-state theorist has to do to make good his claims. He has to specify a physical-chemical state such that *any* organism (not just a mammal) is in pain if and only if (a) it possesses a brain of a suitable physical-chemical structure; and (b) its brain is in that physical-chemical state. This means that the physical-chemical state in question must be a possible state of a mammalian brain, a reptilian brain, a mollusc's brain (octopuses are molusca, and certainly feel pain), etc. At the same time, it must *not* be a possible (physically possible) state of the brain of any physically possible creature that cannot feel pain. Even if such a state can be found, it must be nomologically certain that it will also be a state of the brain of any extra-terrestrial life that may be found that will be capable of feeling pain before we can even entertain the supposition that it may *be* pain. 43

_

⁴³ Hilary Putnam, "The Nature of Mental States." *Philosophy of Mind: Classical and Contemporary Readings*. ed. by David Chalmers. (Oxford University Press, 2002), 77.

Putnam used multiple realizability to argue against the identification of mental and physical properties. If 'pain = C-fiber firing' is necessarily true, pain is realized if and only if C-fiber firing occurs. But we commonly believe that there are other terrestrial creatures able to experience pain and would continue to believe that they experience pain even if we learned they had no C-fibers. Dogs, for instance, feel pain. If we were to learn that the brains of dogs lack C-fibers, we would regard it as more reasonable to reject the claim that pain = C-fiber firing than to reject the claim that dogs feel pain. The identity theory tells us that the mental state-physical state relation is the one-to-one relation of identity, but Putnam convinced many philosophers that it is more reasonable to believe that mental state-physical state relations are one-to-many relation.

Besides Putnam, David Lewis recognized the inadequacies of a simple identity theory. His argument starts with two imaginary men, a madman and a Martian. Both the madman and the Martian can feel pain, but their pains are different from ours and from each others in important respects. The madman's pain has the same physical basis as our pain, but totally differs from ours in its causes and effects. The madman's pain is caused when he has an empty stomach and his feeling of pain causes him to concentrate on mathematics. In fact, the madman's pain has none of the typical causes and effects of normal human pain. He Martian is not human and so the physical basis of his pain is different from ours, namely, the inflation of many smallish cavities in his feet. But the Martian's pain is caused by the sorts of things that cause us pain and, by and large, the

-

⁴⁴ David Lewis, "Mad Pain and Martian Pain." *The Nature of Mind*. ed. by David Rosenthal. (Oxford University Press, 1991), 229.

⁴⁵ David Lewis, "Mad Pain and Martian Pain." *The Nature of Mind.* ed. by David Rosenthal. (Oxford University Press, 1991), 229.

Martian reacts in the same way as we do. So Lewis said that we need a mixed theory, according to which both the Martian and the madman are in pain. By "mixed," he means that although mental states are in fact nothing but physical states, mental *properties* are defined by their causal relations. Lewis said,

My argument is this: The definitive characteristic of any (sort of) experience as such is its causal role, its syndrome of most typical causes and effects. But we materialists believe that these causal roles which belong by analytic necessity to experiences belong in fact to certain physical states. Since those physical states possess the definitive characteristics of experience, they must be experiences.⁴⁶

Lewis' identification of a mental property with its causal role is compatible with multiple realizability, because the same causal role can be performed by different physical states. At the same time, it also explains the contingency when we conceive of mental properties and physical properties. Lewis still believed that metaphysically mental states are nothing but physical states, but a mental property is not identical with any physical property. Mental properties are typically defined *a priori* by their causal relations to other mental properties, stimuli, and behaviors. In other words, on Lewis' account, "pain" is a disguised definite description, a non-rigid designator, that picks out different physical states from species to species. The madman is in pain because he is in the physical state that plays the causal role associated with pain in most humans, even though in his particular psychological economy, that physical state plays a different causal role. The Martian is in pain because he is in a physical state that plays the causal role associated with pain, even though no human could ever be in that physical state. The idea of "pain"

⁴⁶ David Lewis, "An Argument for the Identity Theory." *The Journal of Philosophy*, Vol. 63, No. 1, (1966): 17.

as a disguised definite description makes the property of 'pain' multiply realizable. The simple mind-brain identity is flawed because it cannot allow for multiple realizability.⁴⁷

-

⁴⁷ Strictly speaking, Lewis' theory of mind-body is compatible with a more sophisticated version of the identity theory, one which restricts identity claims to specific populations or species. This possibility complicates the multiple realizability argument against the identity theory, but I think that the argument could be reformulated in a way that overcomes this complication.

CHAPTER 4

FUNCTIONAL REDUCTION

4.1 Varieties of Functionalism

Functionalism has been characterized both as a type of mind-brain identity theory and as a type of property dualism. The method of functional reduction is to functionally characterize the property to be reduced and then find a physical property that fills the functional role. For example, we understand a wine bottle opener is a device for removing corks from wine bottles. All objects which serve this function are wine bottle openers. So functionalists say that 'a wine bottle opener' is a functional property because we can define what it is by its function. If we take the functional description as a non-rigid designator that picks out different physical properties in different contexts, functionalism seems to be a variety of the identity theory. But, if we take the functional description to characterize a higher-level property, functionalism seems to be a type of property dualism. Because functional reduction has such confusing disagreements, some categorization is needed.

Following Ned Block, we can categorize functional theories in terms of whether proposed functional reductions are a part of *a priori* psychology or empirical psychology. The former view is 'analytic functionalism' and the latter is 'psychofunctionalism.' Analytic functionalists such as David Lewis take functional analyses of mental properties as partial analyses of their ordinary meanings. This theory

⁴⁸ Ned Block, "Troubles with Functionalism." *Readings in Philosophy and Cognitive Science*. ed. by Alvin Goldman, (The MIT Press, 1993), 235.

is different from the traditional mind-brain identity theory because it acknowledges that a term like 'pain' in the language of folk psychology is not equivalent to any description expressed in the language of neurophysiology. If there is any *causal/functional* role that is part of the commonsense meaning of 'pain,' then 'pain' can be simply identified with anything which plays that *causal/functional* role. In contrast to analytic functionalism, psychofunctionalists such as Fodor argue that our best empirical theories in psychology will provide functional characterizations for mental properties. Such theories are not reducible and, consequently the mental/functional properties employed by these theories cannot be reduced to the properties of any more fundamental physical theory. Thus, mental properties are a type of functional properties, and both kinds of properties are higher-level properties.

The main difference between analytic functionalism and psychofunctionalism is that the latter emphasizes scientific investigations in the determination of the functional character of mental properties, and these functional identifications are *a posteriori*. From this perspective, psychofunctionalism is incompatible with a mind-brain identity theory because it claims that each mental property is a functional property and not identical with any particular physical property. On the contrary, analytic functionalists claim that a mental property is *the physical property* that occupies a functional role. If scientific evidence makes it plausible to suppose that for a given population of individuals, only one type of physical property in the brain occupies that functional role, then *for that population*, the mental property is identified with that type of the corresponding brain property. Thus, a mental property of individuals belonging to some given population is *identical with* a functionally specified brain property of individuals in the same

population. Consequently, functionalism can be thought of either as a restricted type of mind-brain identity theory or as a type of property dualism. This ambiguity is manifest in the functional role theory of David Lewis.

David Lewis, believes that 'pain' can be defined partially on the basis of an a priori analysis of the causal relations associated with being in pain. We can imagine that a Martian's pain is caused by the sorts of things that cause us pain and Martians react to pain in the same way that we do, but pain in Martians is realized by different physical properties than pain in humans. Lewis argues that what is common to Martian pain and our pain—what makes them both pain—is not anything physical, but rather the fact that they are both instantiations of the same causal relations. To account for the possibility of Martian pain, Lewis says, we need a theory that specifies pain not in terms of its physical properties but in terms of its causal relations. However, for members of the same population such as humans, pain is identical with C-fiber firing, the physical property which typically plays the causal role of pain in normal individuals. We can imagine that a human madman whose C-fibers are firing is in pain even though his state of pain does not play the causal role which it typically does for normal individuals. For the madman, Lewis claims that we need a theory that specifies pain in terms of its physical properties, not in terms of its causal role. The upshot is that Lewis' view is a mixed theory, one that uses a functional analysis to characterize pain across different populations, but identity to characterize pain within a single population.

Other varieties of functionalism do not share the mixed character of Lewis' view. Jaegwon Kim takes functionalism to require the identification of a mental property with a causal/functional property. In *Physicalism, or Something Near Enough*, he argues for

functional reduction. The first step in his model is to functionally analyze the property to be reduced. For example, the property of a gene might be defined in terms of its role in encoding and transmitting genetic information. The second step is to find a realizer that fills the functional role that has been defined. So we scientifically discover that DNA molecules encode and transmit genetic information. The final step is to provide an explanation of how the realizer accomplishes the functional role. Thus, scientists come out genetics to explain how the DNA molecules actually perform encoding and transmitting. 49 When Kim says that a property is a realizer of a functional role, he, of course, assumes this realizer will be a physical property. So, it may seem that Kim is arguing that a mental property is identical with its physical realizer (a causal occupant). But it turns out that he is not saying that a mental property is identical with a physical property which plays the functional role. To return to Lewis's example, although Kim would agree that the Martian is in pain regardless of what specific physical state plays the causal role of pain for Martians, Kim would deny that the madman is in pain. The madman is not in any state that plays the causal role of pain. For Kim, mental properties are necessarily functional properties; no state, including the madman's C-fiber firing, can be pain unless it serves the appropriate causal/functional role.

The difference between Lewis and Kim is that Lewis identifies mental properties with physical properties *for a population* through a functional analysis of our ordinary concept of pain, whereas Kim says that mental properties have necessary connections to their functional roles. Lewis claims that once we define a mental property by its causal

⁴⁹ Jaegwon Kim, Physicalism, or Something Near Enough. (Princeton University Press, 2005), 101-2.

relations and discover a physical property for a population that serves that causal role, having that physical property is sufficient for having the mental property for that population even though there might be exceptions. This allows him to include the madman's pain. But for Kim, a mental property has a necessary connection to the functional property that provides its analysis. Only those physical properties that *actually* serve the appropriate functions can be identified as the realizers of mental properties. Thus, the madman is not in pain because his C-fiber firing does not serve the function of a typical human pain.

Finally, here is an ordinary example to illustrate the difference between the functionalism of Lewis and the functionalism of Kim. Suppose that I bought a package of sticks that have the size and shape of chopsticks. I use almost all of those sticks as chopsticks, but I save some and give them to American children to play with. They have no idea what chopsticks are so they never use them as eating utensils. Are those sticks I give to American children chopsticks? Lewis would say that they are still chopsticks even thought they do not serve the function of chopsticks. They are because they are part of the same package of sticks the majority of which function as chopsticks. The property of being chopsticks is identical with the sticks of that size and shape from that package which typically serves the function of chopsticks. However, Kim says that those sticks American children play with are NOT chopsticks even though the ones I use are chopsticks. The property of chopsticks is identical with the function of chopsticks. In order to be chopsticks, those sticks American children play with must serve the function of chopsticks, because serving that function is what makes them chopsticks. If they never serve the function, they are NOT chopsticks. As for the ones I use, they are used as

chopsticks. They serve the function of chopsticks. Thus, Kim says that they ARE chopsticks.

Determining exactly how functionalism should be characterized is not particularly important in the context of my dissertation. My main concern is to bring into sharper focus that mental properties are analyzed functionally in terms of their causal role and then examine whether functionalism can account for what we take to be the causal efficacy of mental properties.

4.2 Property Dualism, Reduction and Reductive Explanation

Kim's functional model identifies mental properties with functional properties, which makes the mind-brain identity theory false. Saying that mental properties are not identical with physical properties suggests that Kim is a property dualist. Although Kim's functional reduction looks like a type of property dualism, we shall see that Kim does not consider himself to be a property dualist. Generally, property dualism differs from substance dualism by saying that there is only one kind of substance, physical substance, but two different sets of properties, mental and physical. Property dualists do not attempt to reduce mental properties to physical properties. For them, mental properties are irreducible.

The failure of the identity theory---the fact that mental properties are distinct from physical properties---creates a gap between these two types of properties. Following

Joseph Levine, philosophers refer to this as "the explanatory gap." The problem of the explanatory gap is just this: why are mental properties and physical properties correlated in certain specific ways and not others? Conscious experience involves 'a point of view' and that kind of personal point of view does not reduce to any scientific account of the brain. For example, Jon says, "I send my girlfriend a bunch of roses because I love her." Given that the identity theory is false, we cannot say that the feeling of love is identical with L-fiber firing even if that feeling were found to be constantly correlated with L-fiber firing. This leaves a gap to explain: why is the feeling of love constantly correlated with L-fiber firing rather than some other neural properties. Alternately, supposing that neural properties cause the instantiation of mental properties: why does L-fiber firing cause the feeling of love rather than the feeling of pain? Responses to the problem of the explanatory gap are of two types: some philosophers believe that the gap can be closed and these philosophers think that it is possible to reductively explain mental properties. Other philosophers believe that the gap cannot be closed and these philosophers deny the possibility of reductive explanations of mental properties.

Although physical properties do not reduce mental properties, this does not imply that they cannot reductively explain mental properties. The idea of 'reductive explanation' is weaker than the idea of 'reduction.' According to David Chalmers, a reductive explanation of a phenomenon is accompanied by some rough-and-ready analysis of the phenomenon in question and need not require a reduction of that

⁵⁰ Joseph Levine, "Materialism and Qualia: The Explanatory Gap," *Pacific Philosophical Quarterly* 64 (1983): 354.

phenomenon.⁵¹ It is a kind of explanation whereby we are able to understand mental properties in terms of physical knowledge, yet such explanation does not undermine the ontological status of mental properties. Kim says,

It [M-phenomenon] apparently remains an entity with a legitimate, independent standing in its own right; it's only that its existence and character has been made intelligible in light of the underlying phenomena and mechanism. There seems no reason to think a reductive explanation of an M-phenomenon carries any commitment, explicit or implicit, to the claim that it is "nothing over and above" the P-phenomena.⁵²

The idea of reductive explanation shows that even if the project of reducing mental properties to physical properties fails; this failure does not preclude a reductive explanation of mental properties in terms of physical properties.

Some philosophers such as Chalmers believe that some mental properties cannot be reductively explained. ⁵³ They deny both that all mental properties can be reduced and that all mental properties can be reductively explained. According to Kim, the latter claim makes them property dualists in a stronger sense than one that results from the simple denial of the identity theory. Kim's own view is that (1) physicalism requires only a reductive explanation of mental properties, not a reduction of the properties themselves; and (2) some mental properties can be reductively explained, but others cannot. Hence, it explains the somewhat mysterious title of his 2005 book, *Physicalism, or Something Near Enough*.

⁵¹ David Chalmers, *The* Conscious *Mind*. (Oxford University Press, 1996), 43.

⁵² Jaegwon Kim, "Reduction and Reductive Explanation: Is One Possible Without the Other?" *Being Reduced: New Essays on Reduction, Explanation, and Causation*. ed. by Jacob Hohwy and Jesper Kallestrup, (Oxford University Press, 2008), 94.

⁵³ David Chalmers, *The Conscious Mind*. (Oxford University Press, 1996).

4.3 Reductive Explanations via Functional/Causal Role

A reductive functional explanation can be understood as a three step process. The first step is to analyze a concept by using a functional/causal definition. The second step is to find a physical realizer to perform the tasks given by the functional definition. The final step is to develop an explanatory theory to explain how the physical realizer performs that task. Reductive explanation within functionalism means that we give an account of more fundamental processes, through which functional properties fall out. 'A wine bottle opener' is a functional property because we define 'a wine bottle opener' as a device for removing corks from wine bottles. A T-shaped device with a twisted metal screw that slides into the cork thus enabling us to pull the cork out is able to perform the function of a wine bottle opener. A wing-style device has a mechanism with a lever and gears. We screw it into the cork, and then push the levers down so to lift the cork out. These two devices both perform the function of opening wine bottles. Although they have different designs and mechanisms, both of them are physical realizers of 'a wine bottle opener.' Thus, we can develop a theory of 'wine bottle openers' which describes devices for removing corks from wine bottles regardless of color, design, material, and mechanism and that theory contains criteria for determining whether something is a wine bottle opener. To reductively explain a mental property on this model, we first define that property using a functional/causal specification. Then we find a "realizer" to perform the causal specification; this explains why a mental property arises from the physical system. Finally, we explain how the physical realizer performs its task; this explains how the mental property's role is accomplished.

According to Kim, functional properties are higher-level properties. A thing instantiates a higher-level property by instantiating a lower-level property which satisfies the requirements implicit in the higher-level property. For example, my shirt has the higher-level property of being my favorite color because it has a color property, whiteness, which meets the requirements for being my favorite color. Functional properties are higher-level properties for which the requirements are occupying a certain functional/causal role. So the functional requirements for being in pain might be being in a state that is caused by tissue damage and, in turn, causes distress. If C-fiber firing satisfies these functional requirements in humans, then an individual instantiates pain by instantiating C-fiber firing.

The difference between reduction and reductive explanation is that reductive explanation provides a *link* between the explanandum at one level and the purported explanans at another level, and this link explains the correlation between the properties of these two levels. Kim and Chalmers agree that reductive explanations are mediated by functional analyses of mental properties. These analyses make it possible to construct an explanatory "bridge" from the lower-level physical facts to the phenomenon in question.⁵⁴ Allow me to use Kim's and Chalmers' strategy to illustrate this:

- (1) Jane is in the neural property N at time t.
- (2) N satisfies the causal role C in human beings.
- (3) Having pain = $_{def.}$ being in some state satisfying C.

⁵⁴ David Chalmers, *The Conscious Mind*. (Oxford University Press, 1996), 44.

(4) Therefore, Jane is in pain at time t.⁵⁵

The third line is a functional definition, a causal specification described in functional vocabulary. This line reduces the property of pain to the causal/functional property given by the description 'C'. The explanatory link between the physical property N and the functional property, pain, is provided by line (2) — a law-like statement. Together line (2) and (3) explain what pain means in terms of the performance of some function by the neural property N at t. Finally, once we have explained how the function is performed by N, we have explained both why and how Jane is in pain at time t. Therefore, the explanatory gap is closed.

Kim believes that functional reductions also enable us to understand the causal power of mental properties. For example, we can explain why Jane winced because of her pain:

- (1) Jane enters state N_1 ; state N_1 brings about state N_2 .
- (2) Finding physical realizers

In humans, N_1 realizes paper cuts.

In humans, P satisfies causal role C where C includes being caused by N_1 and causing N_2 .

In humans, N₂ realizes wincing.

(3) Functional definitions

Paper-cut = $_{def.}$ razor-like cut in skin caused by paper and causing pain.

68

⁵⁵ I am modeling this example by following Kim's on *Physicalism, or Something Near Enough*. (Princeton University Press, 2005), 111.

Pain = $_{def.}$ being in some state satisfying C.

Wince = def_facial expression caused by pain.

(4) Therefore, Jane winced because of the pain brought about by her paper cut. The paper-cut pain is defined at (3) by its causal relationships, so it presumably has the causal efficacy based on the functional analysis. The physical/causal processes are governed by law-like statements given in (2). The explanatory link between this type of mental causation and its corresponding neural properties is provided by lines (2) and (3). The whole account allows us to explain why Jane winced *because* of her pain. The ability of functional reductions to explain the causal powers of mental properties relies according to Kim, on the Causal Inheritance Principle. He characterizes this principle as follows,

If a second-order property F is realized on a given occasion by a first-order property H (that is, if F is instantiated on a given occasion in virtue of the fact that one of its realizers, H, is instantiated on that occasion), then the causal powers of this particular instance of F are identical with (or are a subset of) the causal powers of H (or of this instance of H). ⁵⁶

According to the Causal Inheritance Principle, every instantiation of a functionalized mental property is identical with an instantiation of the physical property which realizes it. The instantiation of the causal links betweens Jane's paper cut, her pain, and her wince is identical with the instantiation of the causal links from her paper cut N_1 , from N_1 to N_2 , and from N_2 to her wince. Thus, mental causation has to be instantiated at the physical level according to Kim.

_

⁵⁶ Jaegwon Kim, *Mind in a Physical World*. (The MIT Press, 2000), 54.

The identity theorists solve the problem of mental causation by identifying mental properties with physical properties and thereby transferring the causal powers of mental properties to the causal powers of physical properties. While Kim rejects the identification of mental and physical properties, he also transfers the causal power of paper-cut pain to the causal powers of the neural properties. According to Kim, pain and mental states generally, have no causal powers of their own. Kim eliminates the causal power of all mental phenomena as much as the identity theorists do. The only difference is that for the identity theorists statements such as, 'N₁ causes N₂' *are rewrites* of ordinary statements such as 'Jane winced because of her pain,' whereas for Kim, statements such as, 'N₁ causes N₂' *help to explain* ordinary statements such as 'Jane winced because of her pain,' This similarity between Kim's account of mental causation and that of identity theorists will be the basis of my criticism of Kim in chapter 5.

4.4 The Limits of Kim's Model of Functionalism

Kim's physicalism has evolved primarily from the effort to reconcile mental causation. In his view, reductive explanation is required to save casual efficacy for mentality, but only some mental properties can be reductively explained. In other words, he believes that mental causation requires physical reduction, but physical reduction is not required for all mental properties. Thus, the reductive physicalism Kim endorses is *conditional* reductionism because it solves the problem of mental causation only. Kim believes that qualitative characters (qualia) are irreducible. He adopts the traditional distinction between two kinds of mental properties. Those Kim calls "intentional/cognitive properties" are properties that play some causal/functional role in

the life of an organism. This kind usually includes beliefs, desires, memories, and perceptions. Intentional/cognitive properties can be reductively explained in the manner discussed in section 4.3. By contrast, "phenomenal/qualitative properties" are those properties of subjective experience with distinctively intrinsic qualitative characters. This kind includes pains, itches, after-images, etc. Phenomenal properties cannot be reductively explained.⁵⁷ Because Kim believes that mental properties have causal powers only if those properties can be reductively explained in virtue of their causal/functional roles, and because he believes that phenomenal properties cannot be reductively explained, phenomenal properties have no causal power.

A subjective experience is *what it is like* within each person. It is a qualitative phenomenon and, according to philosophers such as Kim and Chalmers, the qualitative properties that constitute *what subjective experience is like* never affect what we say, how we behave, or any bodily property we possess. In other words, although phenomenal properties exist within each person, their distinctive character is completely inaccessible from a 3rd person point of view. The most famous argument for the essential privacy of phenomenal properties is Chalmers' zombie example. Zombies are creatures physically identical with us but lacking conscious experience. If these creatures function physically just like us but have no conscious experience, then conscious experience does not play any role in our behavioral system. Because zombies are physically/behaviorally identical with us, the presence or absence of qualitative experiences cannot be determined through any objective or scientific investigation. If phenomenal properties performed any

⁵⁷ Jaegwon Kim, *Physicalism, or Something Near Enough*. (Princeton University Press, 2005), 162.

causal/functional work, they would be detectable. But they are not detectable. Thus, Kim argues that phenomenal properties are causally inert. Only intentional/cognitive properties have causal power. Thus, Kim's physicalism is not complete, because it reductively explains only intentional/cognitive properties. Phenomenal properties remain unreduced and epiphenomenal.

Contrary to Chalmers and Kim, Michael Tye has a different account of phenomenal properties. He thinks that phenomenal properties have external relationships to the physical world. Phenomenal properties, seeing the color red for example, have non-conceptual representational contents and those contents are associated with phenomenal knowledge. He thus argues that phenomenal properties can be explained in terms of their distinctive functional roles.

Tye argues that we can explain the phenomenal concept of color experience, for example, without admitting any inaccessible subjective qualities of the experience. Frank Jackson's case of Mary illustrates that phenomenal concepts are associated with phenomenal knowledge. Mary was trapped in a black-and-white room and has been there since she was born. She was a color scientist so she possessed of all the physical/neural information about color and color vision. But she never had any experience of red, green, etc. while she remained in the room. One day, she left the room and saw red roses. Upon seeing the roses, Mary had her first experience of red and thereby realized the distinctively qualitative character associated with an experience of red. According to Frank Jackson, Mary came to *know* something she did not know---what it is like to

⁵⁸ Jaegwon Kim, *Physicalism, or Something Near Enough*, (Princeton University Press, 2005), 169-70.

experience red. Jackson concluded that what Mary came to know was something over and above any physical/functional knowledge of the brain she had. Tye, by contrast, agrees with Jackson that Mary discovered something new when she first experiences red, but Tye denies that she came to know a non-physical fact; indeed he denies that she came to know a new fact at all. Tye argues that there is a difference between factual knowledge and knowledge by acquaintance. When Mary experiences red for the first time, she learns no new facts, but she becomes acquainted with red.

Mary in her room knows all the physical facts *about* the subjective character of experience of red. But there is a perfectly ordinary sense of 'know' under which she does not know the thing that is the subjective character of the experience of red. She is not acquainted with that thing. When she leaves the room and become acquainted with the phenomenal or subjective character of the experience of red, thereby she knows it. This is genuinely new knowledge, logically distinct from her earlier factual knowledge. ⁵⁹

Before Mary left the room, she had exhaustive factual knowledge about red, but she was not acquainted with red. Her situation was analogous to someone who had extensive knowledge about President Obama, but who had never met the man. Seeing red enables Mary to have a kind of knowledge that no amount of information about 'red' can supply. Therefore, Tye concludes that qualitative properties in fact have external relationships with physical information.

On Tye's view, the apprehension of phenomenal/qualitative properties is not subjective and private, because qualia are relational properties. The awareness/introspection of qualitative properties delivers a connection to physical objects: to external objects in the case of sensory awareness, or to my body in the case of

-

⁵⁹ Michael Tye, *Consciousness Revisited*. (The MIT Press, 2009), 131-32.

introspection. Therefore, the introspection/awareness of *what it is like* is an epistemic state whose character is independent of the experience.

The phenomenal character of an experience, then, is out there in the world (or in the body, in the case of bodily sensations). It is not a property of the experience at all. It is a complex of properties *represented* by the experience. In being aware of the external qualities, we are aware of phenomenal character.⁶⁰

The content of phenomenal/qualitative properties is in fact a type of intentional content and, like intentional content generally, stands in external relationships to physical facts. Simply speaking, although what-it-is-like-to-experience-red is private to Mary, the redness of that experience is a relational fact about it. It refers to the external property of redness.

On this interpretation, subjective qualities are not inaccessible. We are aware of what we are experiencing and our awareness makes subjective qualities map to our knowledge of them. If Tye's view is correct, it might open the way for the functional reduction of phenomenal properties, and thereby provide an account of their causal efficacy. Unfortunately, Tye does not discuss the issue of mental causation extensively and although Tye's view on phenomenal properties is interesting, discussing its pros and cons is beyond the scope of this present work. I hope to take up this issue in future research.

_

⁶⁰ Michael Tye, Consciousness Revisited. (The MIT Press, 2009), 119.

4.5 Another Reductionist Model---Bridge-Law Reduction

Before functionalism became popular, many philosophers believed that there was another way to reductively explain the correlation between mental and physical properties. If there were psycho-physical laws between the mental and the physical, then mental properties would be reductively explained in terms of more fundamental vocabularies. This model of reduction allows us to derive the laws of psychology from the laws of physics via the psycho-physical laws. In this context, those psycho-physical laws are "bridge laws" connecting the two levels.

Bridge-law reduction is designed for a reduction of *theories*. It closes the explanatory gap because laws of psychology, together with psychological properties referred to in such laws are explained by being deduced from physical laws. The psychophysical laws provide a condition to connect the psychological terms and physical terms. Following bridge-law reduction's doctrine, the explanation of the correlation of 'pain' and 'C-fiber firing' is as follows,

- (1) Jane's C-fiber is firing at time t.
- (2) Bridge law L_{pain}: For humans, pain occurs if and only if C-fiber firing occurs.
- (3) Therefore, Jane is in pain at time t.

The second line is a scientific discovery that allows us to derive psychological facts from the laws of a base theory; thus (2) is an empirical and contingent law. Unlike the mindbrain identity theory, the connection between mental properties and physical properties is contingent, not necessary. Unlike functionalism, bridge-law reduction does not appeal to any type of analysis of mental properties. According to defenders of bridge law reduction,

L_{pain} is sufficient to explain correlation of 'pain' and 'C-fiber firing' and to close the explanatory gap.

Kim says that while bridge laws do enable us to derive psychological laws from physical laws, they do not provide genuinely reductive explanations.

For let L be a law of the theory to be reduced; we can use the bridge laws as definitions to rewrite L in the vocabulary of the base theory, turning L into L*, a statement in the language of the base theory. Now, either L* is derivable from the laws of the base theory or it is not. If it is, the derivability condition is met. If it is not, add L* to the base theory as an additional law. This is justified, since L* is expressed purely in the predicates of the base theory and, in missing L*, the original base theory is missing a general truth within its domain.

According to Kim, bridge laws are themselves "unexplained auxiliary premises" in the above derivation and are themselves in need of explanation. Since bridge laws must be added to the laws of the reducing theory to make the derivation possible, there is actually a price to be paid. To derive a law of psychology (M) from a law of the base theory (P) with the help of bridge laws, we need the conjunction (P & bridge laws) as a premise. Because this is the case, we can no longer say that the base theory is P; instead, it becomes (P & bridge laws), which is a hybrid theory containing both the base theory and the bridge laws. When bridge laws become auxiliary premises of the theory P, the base theory changes, and we can no longer say that M has been reduced to the *base* theory, P. We must say instead that M has been reduced to a hybrid theory, P & bridge laws.

If bridge laws are supplementary premises of P, we in fact extend the predicates of P by adding these laws. The psycho-physical laws, "bridge laws," were supposed to

⁶¹ Jaegwon Kim, *Physicalism, or Something Near Enough*. (Princeton University Press, 2005), 99.

⁶² Jaegwon Kim, Mind in the Physical Word. (The MIT Press, 2000), 96.

close the gap between psychology and physics. But when the hybrid theory becomes the new base theory, bridge laws no longer provide a connection between the vocabulary of psychology and that of a base theory. Rather, they become laws of the base theory itself. Most importantly, the bridge laws themselves stand in need of explanation. Traditionally, explanations of the non-basic laws of a theory took the form of derivation of those laws from more basic laws of that theory. But bridge laws cannot be derived from the basic laws of the reducing theory. This produces a new explanatory gap between these particular non-basic laws of the reducing theory and the more basic laws of that theory. The emergence of this new explanatory gaps shows that the proposed reductive explanation was a sham.

I think that Kim is indeed right in criticizing bridge-law reduction on this point. Bridge-law reduction simply assumes the integrity of what exactly needs to derived and explained. However, I think that Kim's model has a component that plays an analogous role to that of the bridge laws. On Kim's account, for a mental property to play a causal/functional role is for it to have a physical realizer that instantiates the causal role definitive of the mental property. That is, the causal role of a mental property is given by a set of causal laws in which the mental property figures. To use an example, in folk psychology M, pain is defined as a state caused by tissue damage and that causes wincing. Let us now ask what is meant by saying that some physical state fills the causal role of 'pain' as defined by folk psychology M. I suggest the explication goes like this: if any person feels pain (say m as a mental property of pain), then his C-fiber fires, where C-fiber firing belongs to a set of physical realizers $(p_1, p_2...p_n)$ of 'pain.' For every person who feels pain $(m_1, m_2...m_n)$, there is a law for each pain, $L_n(m_1, m_2...m_n)$, of

psychology M, and there is a physical law for every physical realizer, $L_n^*(p_1, p_2, ...p_n)$, of 'pain' of a base theory T*. If we say that p_i occupies the m_i 's causal role, there is an approximate mapping from one law (L_i) of psychology (M) to one law (L_i^*) of a base theory (T^*) . For example, we may say that C-fiber firing fills the causal role of pain in a sense that for folk-psychological laws saying that tissue damage causes pain and pain causes winces, there are physical laws saying that tissue damage causes C-fiber firing and C-fiber firing causes winces.

On this account of "causal role filling," law L_i of psychology M is derivable from law L_i* of the base theory T*. A physical realizer p_i of the mental property m_i is identical with m_i itself in virtue of p_i occupying in L_i* the position m_i occupies in L_i. It seems the account of "causal role filling" can be seen as a device for linking certain predicates of L_i* with corresponding predicates of L_i. P_i fills the causal role of m_i only if each of the other physical realizers $(p_1...p_n)$ in L_i^* can be similarly linked to (or even identified with) a corresponding mental property $(m_1...m_n)$ in L_i . So, the account of "causal role filling" under Kim's functional model plays the equivalent role as bridge laws. As bridge-law reduction leaves a new explanatory gap unexplained, Kim's model encounters a similar difficulty. "Causal role filling" is the additional premise of the base theory because this law tells us that C-fiber firing fills the causal role of pain. The base theory and this additional premise together expand the original base theory and the new base theory connects physical realizers with functional properties. Property identification, as we say from the mind-brain identity theory, does not close the gap because it simply replaces functional properties with physical realizers. Furthermore, Kim's functionalism works only for those mental properties which are functionally definable. Phenomenal properties

are not functionally definable; thus, there is no explanation for phenomenal properties from his functionalism. Kim's functionalism leaves phenomenal properties unexplained.

Where does this leave us with respect to the issue of psycho-physical reduction? I have explained Kim's response to the identity theory and shown that we can reductively explain mental properties without ontologically reducing those properties as the identity theory does. There is no need to pursue a stronger reductionist model. However, while Kim criticizes bridge-law reduction as a failed reductionist model because it lacks the ability to close the explanatory gap, I think that there is a good reason to resist Kim's contention that bridge-law reduction is "the wrong battlefield on which to contest the issue of reduction." As I argued, if bridge-law reduction cannot provide reductive explanations, neither can Kim's own functional model. The account of "causal role filling" implies a mapping relation whose job is working as a bridge law. If bridge-law reduction does not close the explanatory gap, neither does Kim's functional reduction.

_

⁶³ Jaegwon Kim, Mind in the Physical Word. (The MIT Press, 2000), 26.

CHAPTER 5

EMERGENTISM

5.1 Non-Reductive Physicalism

Many philosophers who have been persuaded by the multiple realization argument embrace non-reductive physicalism. They sought in the Supervenience Thesis a satisfying metaphysical statement in physicalism without reductionism. A trend of non-reductive physicalism seems to be in place. In contrast to reductive physicalism, non-reductive physicalism says that while mental properties are fundamentally physical, they are not reducible to any physical properties. But as we saw in chapter 4, philosophers such as Kim still believe that, through causal/functional analyses, mental properties can be reduced to functional properties and that these functional properties are, in turn, realized by physical properties. In this way, functionalists propose to reductively explain both the presence of mental properties and their causal powers. Kim has developed an argument arguing that non-reductive physicalism is ultimately untenable because it cannot accommodate the causal efficacy of mental properties. The heart of the non-reductive physicalists claim is that mental properties have causal power but it cannot be reductively explained as functionalists have proposed.

Most non-reductive physicalists will agree that the following three theses. ⁶⁴

(1) Mental properties strongly supervene on physical/biological properties. That is, if any system s instantiates a mental property M at t, there necessarily

80

⁶⁴ Jaegwon Kim, *Physicalism, or Something Near Enough*. (Princeton University Press, 2005), 33-5.

- exists a physical property P such that s instantiates P at t, and necessarily anything instantiating P at any time instantiates M at that time.
- (2) Mental properties are not reducible to, and are not identical with, physical properties.
- (3) Mental properties have causal efficacy---that is, their instantiations can, and do, cause other properties, both mental and physical, to be instantiated.

Non-reductive physicalists believe that the reality of mental properties is shown in their causal influence in the physical world even though they are not physically reducible. Kim argues that any type of physicalism worth the name must maintain that ultimately all causation is physical causation. This is why, according to functionalism, mental properties must be realized by physical properties. Thus, to be proper physicalists, Kim says that we have to accept the Physical Realization Thesis:

(4) All mental properties are physically realized; that is, whenever an organism, or system, instantiates a mental property M. it has some physical property P such that P realizes M in organism of its kind.⁶⁵

So, functionalists use 'realization' to help to close the explanatory gap. The fact of realization provides an explanatory connection between the causal power of mental properties and the causal power of physical properties. This adoes not imply that mental properties are identical to physical properties. It just brings out a dependent relationship which preserves a limited form of the autonomy of the mental. Along with the Physical

_

⁶⁵ Jaegwon Kim, "The Nonreductivist's Trouble with Mental Causation." *Mental Causation*. ed. by John Heil and Alfred Mele. (Oxford University Press, 1993b), 198.

Realization Thesis, Kim also believes that physicalists have to commit to the Principle of Causal Closure of Physical Domain.

(5) If a physical event has a cause that occurs at t, it has a physical cause that occurs at t. 66

This principle usually goes with the Principle of Causal Exclusion:

(6) No single physical event can have more than one sufficient cause at any given time--- unless it is a genuine case of causal over-determination.⁶⁷

They together claim that no mental causes are rquired to account for the occurrence of any physical effect. The physics is causally closure and mental causes have to be physically realized.

As physicalists, Kim argues that non-reductive physicalists, must accept theses (4), (5), and (6). But as non-reductive physicalists, they also accept thesis (3). Thus, Kim argues that their position is inconsistent. What is striking about these theses is that they lead to 'downward causation,' saying that mentality takes on a causal life of its own 'downward' to affect what goes on in the underlying physical processes.

If mental properties have causal power in their own right, there must be an appropriate causal law that invokes the instantiation mental properties as sufficient conditions for other effects. There are two possibilities, either these effects are other mental events or they are physical events. The former case is mental-to-mental causation (M causes M*). The latter case is mental-to-physical causation (M causes P*).

According to the Physical Realization Thesis, the mental effect M* is physically realized

⁶⁷ Jaegwon Kim, *Physicalism, or Something Near Enough*. (Princeton University Press, 2005), 42.

82

⁶⁶ Jaegwon Kim, *Physicalism, or Something Near Enough*. (Princeton University Press, 2005), 43.

by P*. If M has a causal efficacy to M*, it must affect M*'s physical realizer P* as well. So mental-to-mental causation (M causes M*) presupposes mental-to-physical causation (M* causes P), which is downward causation.

Kim's main purpose is to argue that downward causation is incoherent, because we cannot have downward causation without committing physical-to-physical causation (P causes P*). We cannot avoid the fact that M, as a mental property, has its own physical realizer, P. If downward causation is possible, why not say that M's causation to P* comes from P's causation to P*? Here we have a more basic causal relation between two physical properties, P and P*. Ultimately, mental-to-mental causation (M causes M*) is grounded in physical-to-physical causation (P causes P*). Mental causation has to be done at the physical level according to (4), (5), and (6). But it is a consequence that non-reductive physicalists seek to avoid because it violates (3) and leaves mental properties causal inert.

Kim argues that the epiphenomenalist consequence regarding mental properties does not easily go away within non-reductive physcialist framework. We see that M and M* are realized by P and P* respectively. The M-to-M* causation can be instantiated by P-to-P* causation; thus, all causal relationships are implemented at the physical level. There is no causal work is left over for M to do if P is a sufficient cause of P*. Mental properties can only bring about physical changes if they are realized in some physical properties. The result is that the purported mental causes are epiphenomenal. So, either non-reductive physicalists must give up their idea of non-reduction which means accept functionalist account of mental causation, or they have to accept that mental properties are epiphenomenal.

5.2 Emergent Properties

Some philosophers define non-reductive physicalism by putting the Physical Realization Thesis aside and replacing it with another concept, emergence. They accept that mental properties supervene on the physical properties and that mental properties and physical properties are correlative, but the correlation is not explained with physical realization. Through emergence, it is not necessary to assume the Principle of Physical Realization. In other words, since physical realization is not the only relation to explain mind-body correlation, there is no need to hold the Physical Realization Thesis. They also believed that emergence is more consistent with the idea of 'non-reduction' than physical realization.

The tradition of emergentism, known as "British Emergentism" began in the middle of 19th century and continued into the early of 20th century. Much of the defense of emergentism in this era was centered on chemistry and biology, such as John Stuart Mill's *System of Logic* (1843), George Henry Lewes's *Problems of Life and Mind* (1875), *Samuel Alexander's Space, Time, and Diety* (1920), Lloyd Morgan's *Emergent Evolution* (1923), and C. D. Broad *The Mind and Its Place in Nature* (1925). Early British Emergentists held an ontology of a layered world. The world is divided into discrete levels, with physics as the base level, followed by chemistry, biology, psychology, and so on. Each level is arranged in terms of organizational complexity at the bottom level. As we move up to each level, science at that level becomes specialized with an increasingly complex structure. The relationship connecting each level is emergence, but its exact nature is left vague. British Emergentists sought to develop a middle way to explain

special sciences employing a physical-chemical mechanism at the base while retaining irreducibly vital qualities or processes at higher levels.

Some forms of supervenience require only a weak link between mental properties and physical properties. Weak supervenience tells us that there can be no difference in the supervening domain without a difference in the supervenient base. For example, there can be no difference in the mental domain without a difference in the underlying physical domain. Any change at the mental level corresponds to a change at the physical level, but not vice versa. Weak supervenience only requires a correlation between mental properties and physical properties, but it does not detail how physical properties determine mental properties. Thus, some non-reductive physicalists add of mind the concept of 'emergence' to rich the concept of supervenience. Emergent properties are supervenient on physical properties, but also "over and above" the underlying physical properties.

Samuel Alexander explains this relationship as follows,

Physical and chemical processes of a certain complexity have the quality of life. The new quality life emerges with this constellation of such processes, and therefore, life is at once a physic-chemical complex and is not merely physical and chemical, for these terms do not sufficiently characterize the new complex which in the course and order of time has been generated out of them. Such is the account to be given of the meaning of quality as such. The higher quality emerges from the lower level of existence and has its roots therein, but it emerges therefrom, and it does not belong to that level, but constitutes its possessor a new order of existent with its special laws of behavior. The existence of emergent qualities thus described is something to be noted, as some would say, under the compulsion of brute empirical fact, or, as I should prefer to say in less harsh terms, to be accepted with the "natural piety" of the investigator. It admits no explanation. 68

⁶⁸ Samuel Alexander, Space, Time, and Deity. Vol.2, (Macmillan and Co, 1920), 46-7.

Although emergence is a relation connecting two levels of reality, emergent properties are novel properties arising from the underlying basic properties. Novelty cannot simply mean "not having instanced previously." Nor can it mean "not had by any of the object's proper parts." Alexander's understanding makes the concept of emergence not a mere "resultant" of underlying physical properties. To say that emergent properties are novel properties, the discussion of 'emergence' is spoken in terms of 'non-structurality.' A property is structural if and only if there is nothing more to having this property than being a structure wholly composed of parts from its underlying physical properties. For example, H₂O is composed of hydrogen and oxygen. The notion of an emergent property can then be understood in part by way of contrast with structural properties. An emergent property is a property of a composite system whose existence cannot be explained in terms of the structures of the parts of the composite system. For example, at one time many scientists and philosophers believed that the transparency of water could not be explained by the composite structure of hydrogen and oxygen. If mental properties were "resultant" of the underlying physical properties, then such mental properties would be nothing over and above structures composed of physical properties and thereby ontologically reducible. By contrast, emergence formulates the relation between mental properties and physical properties in such a way that mentality arises from the physical, but does so in a way that makes it genuinely novel. My central purpose in what follows is not to explain how British Emergentism worked in detail, but to explain why it has fallen.

British Emergentists maintained that everything is made of physical particles.

There is a hierarchy of levels of organizational complexity of physical particles, including chemistry, biology, and psychology. So each level in the hierarchy corresponds

to a certain science and each higher level science is composed by but not reduced to the kinds of lower level. Physics is at the base of the hierarchy of sciences in that all higher-level properties emerge from the organizational complexity of physical particles. The most notable example used by British Emergentists was that through physical science, we know that water is composed of hydrogen and oxygen and that salt is composed of sodium chloride. However, British Emergentists thought that physical science could supply no explanation of the fact that an aggregate of H₂O molecules has the power to dissolve an aggregate of NaCl (sodium chloride's formula). Such higher-level laws must be accepted with Samuel Alexander's natural piety. From water's formula, H₂O, we cannot predict the dissolution of NaCl. Whether the dissolution of NaCl in water can be explained or reduced to lower-level explanations is an *a posterior* issue for British Emergentists. When higher-level phenomena are able to be explained by lower-level knowledge, the higher-level phenomena are non-emergent.

British Emergentists were well-aware that the entities in the domains of the special sciences are different kinds of aggregates of particles. Higher laws cannot be explained from the lower levels. Although higher laws constitute novel relationships, they depend on the aggregation of lower-level particles. For British Emergentists, everything is made of physical particles so new, emergent characteristics are not characteristics of distinct or even an emergent substances. In this sense, the notion of 'emergence' is not ontological. When they spoke of an emergent property, they were referring to a special feature that is unable to be *explained* by lower-level laws. Finally, British Emergentists believed that these higher-level features had causal powers whereby

they could affect the processes occurring at lower levels. In a nutshell, British Emergentists define an emergent property as:

- (1) An emergent property supervenes on its underlying property.
- (2) An emergent property cannot be reduced to or reductively explained in terms of its underlying properties.
- (3) An emergent property can directly influence the pattern of behavior involving its underlying properties, and this influence is unpredictable and unexplainable from the laws of underlying system.

Among these basic characteristics of emergence, we must emphasize the third more than the rest. If an emergent property is a novel property arising from its underlying system, it would be impossible for an observer having a completely adequate understanding of the patterns at an underlying system to predict its emergent properties and behaviors.

Therefore, if we embrace emergence, we have to embrace the notions of irreducibility and unpredictability.

So far as I can tell, the fall of British Emergentism occurred because it did not enable scientists to predict what would happen in the higher-level sciences. British Emergentism says whether or not a property at a higher level is an emergent property is an *a posteriori* issue. The discovery of emergent properties is an empirical fact, and their emergence has no explanation. Thus, British Emergentism did not guarantee any explanation for how higher-level laws are derived from the lower-level laws. Emergent properties are not predictable on the basis of the lower-level system which determines the instantiation of lower-level properties. British Emergentists simply said that there are emergent properties and laws and they have emerged from the underlying fundamental

system, but we have no explanation of how such emergence occurs in special sciences with a natural piety.

5.3 Mental Causation in a Physical World

If our world is fundamentally physical, how can mentality make any difference? If the doctrines of British Emergentism are correct, we have to assume that such an answer should be compatible with the idea that mentality is hierarchically reduced to physics but the functions of mentality are irreducible. But British Emergentists admitted that current science lacks the explanation of emergent properties because of limits on current scientific knowledge. However, as our scientific knowledge grows, we might discover emergent properties in the future. I think that Kim's functionalism partly agrees with British Emergentism, because they both make the scientific discovery of higher-level properties a matter of scientific investigations. Kim argues that the Supervenience Thesis determines the physical priority of emergent properties. Every emergent property has its own physical base. He had used the supervenience argument to argue that the higher-level causation (mental causation) has to be instantiated at the physical level if the Supervenience Thesis is one of the basic theses of non-reductive physicalism. The supervenience argument is as follows: ⁶⁹

89

(i) Either mind-body supervenience holds or it fails.

-

 $^{^{69}}$ Jaegwon Kim, $\it Mind~in~a~Physical~World.$ (The MIT Press, 2000), 39-46.

- (ii) If mind-body supervenience fails, there is no visible way of understanding the possibility of mental causation.
- (iii) Suppose that an instance of mental property M causes another mental property M* to be instantiated.
- (iv) M* has a physical supervenience base P*.
- (v) M* is instantiated on this occasion: (a) because, ex hypothesi, M causes M* to be instantiated; (b) because P*, the physical supervenience base of M*, is instantiated on this occasion.
- (vi) M caused M* by causing P*. That is how this instance of M caused M* to be instantiated on this occasion.
- (vii) M itself has a physical supervenience base P.
- (viii) P caused P*, and M supervenens on P and M* supervenes on P*.
- (ix) The M-to-M* and M-to-P* causal relations are only apparent, arising out of a genuine causal processes from P to P*.
- (x) If mind-body supervenience fails, mental causation is unintelligible; if it holds, mental causation is again unintelligible.

Kim says that the failure of mind-body supervenience would require a violation of the commitment of the Principle of the Causal Closure of Physical Domain. Why? Kim answers,

If you pick any physical event and trace out its causal ancestry or posterity, that will never take you outside the physical domain. That is, no causal chain will ever cross the boundary between the physical and the

nonphysical...If you reject this principle, you are ipso facto rejecting the in-principle completeability of physics.⁷⁰

I think that Kim confuses the Principle of the Causal Closure of the Physical Domain with the Principle of the Completeness of Physics. The Principle of the Completeness of Physics is the claim that every physical effect has a sufficient physical cause and no effect has more than one distinct sufficient cause unless it is a genuine case of overdetermination. So the idea that no causal chain will ever cross the boundary between the physical and the nonphysical is described in the Principle of Completeness of Physics. However, tracing the causal ancestry of a physical event indeed never takes us outside the physical domain, but it might have non-physical cause if we let it. Boundary-crossing causal chains are not excluded by the Principle of the Completeness of Physics, but they are compatible with the Principle of the Causal Closure of the Physical Domain. Not being as strong as the completeness of physics, the Principle of the Causal Closure of the Physical Domain says that if a physical effect has a cause, it has a physical cause. It is compatible with this idea that a physical effect is able to have a mental cause as long as it accompanies a physical cause.

In order to avoid a dualist view, Kim intends to explain away mental-to-mental causation in terms of physical causation. This takes an intuition, if M causes M*, and according to the supervenience thesis, M* supervenes on P*, then M must also be a cause of P*. It tells us that mental-to-mental causation in some sense is instantiated by the mental-to-physical causation. Now, we have M's and M*'s supervenience bases, P and P*. If M has a causal effect on P*, P as being the supervenience base of M should have

⁷⁰ Jaegwon Kim, *Mind in a Physical World*. (The MIT Press, 2000), 40.

the same causal effect on P* as well. Kim claims that we have to find reasons for "taking P as preempting the claim of M as a cause of P*". M and P are both possible causes of P*, but Kim believes that this situation is not a genuine case of over-determination. M supervenes on P and P inherits M's causal power, and thus if P causes P*, there is no any causal work left for M. But, what is wrong with the view that a physical event is the result of a physical event and a mental event acting jointly? When the mind-body supervenience works, all mental causes are dependent but not reduced to physical causes. Kim simply says, "It is difficult to see how M and P together can pack any more causal power than M alone or P alone." Why difficult? I assume that Kim simply applies the Principle of Causal Exclusion and the Principle of Causal Closure of Physical Domain again.

Latter in his 2005 book, *Physicalism, or Something Near Enough*, Kim argues that mental causation has to be instantiated by physical causation through the help of functional definition. Kim obviously is not willing to give up the completeness of physics and that is the reason that he insists on both the Principle of the Causal Closure of the Physical Domain and the Principle of Causal Exclusion to exclude mental causes. I am not saying that the above two principles are false. The point I want to make is that emergentism is compatible with all theses of non-reductive physicalism except the Physical Realization Thesis. The concept of emergence need not presuppose physical realization. The consequence of "mental properties emerge from physical properties" is that physical properties are necessary conditions for mental properties but without

⁷¹ Jaegwon Kim, *Mind in a Physical World*. (The MIT Press, 2000), 43.

⁷² Jaegwon Kim, *Mind in a Physical World*. (The MIT Press, 2000), 44.

emergent laws, they are not sufficient. Thus, we have no solid reason to believe that we have to accept the completeness of physics if we accept the supervenience thesis. A person supporting the idea of mental causation means that he agrees that mental properties intervene somehow in physical processes to produce a result. In the following section, I would like to provide a new concept of emergence, that is consistent with the theses of the irreducibility of the mental, mind-body supervenience, mental causal efficacy, the Principle of the Causal Closure of the Physical Domain, and the Principle of Causal Exclusion.

5.4 A New Concept of Emergence

Mental causation requires a more metaphysically robust version of emergence than permitted by the British Emergentists' view of emergence. If we accept that emergence serves only an epistemological function, Kim's supervenience argument could be right: there is no genuinely mental causation. For British Emergentists, mental causation is a merely a placeholder for some yet to be discovered physical causal account. If the concept of 'emergence' has only this epistemological function, it does not introduce a type of mind-body relationship significantly different from other reductionist accounts. This concept of 'emergence' cannot offer any more help to mental causation than reduction. Therefore, we need to think of emergent properties not just as a sort of structural property, but as something that is truly "over and above" those aggregative particles. I think that emergence must embrace the idea of ontological novelty. This means that irreducible higher-level properties do arise out of complex configuration of underlying lower-level properties. In order to defend this ontological brand of emergence,

I argue that the relationship between higher-level properties and lower-level properties is a kind of nomological relationship instead of structural composition. In contrast with the British Emergentists' concept of emergence, I want to argue that a new concept of emergence should be understood based on the concepts of novelty, nomological emergence, and non-structurality.

Non-reductionists share a concept of mind whereby our mentality is not merely a process of neural transmissions. Instead, mind is understood as the natural result of processes involving neurons. However, our current scientific knowledge suggests that the world is fundamentally physical and integrated. As long as the doctrine of ontological physicalism that is, the doctrine that all individual things are physical things, is maintained, the difficulty threatening mental causation remains. Is mental causation compatible with ontological physicalism and with the possibility of ontological novelty? If the answer is "Yes," the novelty of emergent properties should be construed as metaphysical and not merely epistemological. One example of ontological emergence might be found in the Big Bang. The Big Bang theory says that the whole universe has expanded from a primordial hot and dense initial state. The theory does not provide any explanation of the initial conditions that preceded this state but describes and explains the general evolution of the universe since the bang. In physical cosmology, nucleosynthesis took place after the Big Bang occurred and it was responsible for the formation of primordial nuclei and that then transformed into protons and neutrons. Suppose that at the most basic level, ontologically speaking, there is quantum flux, an undifferentiated domain of reality not unlike what William James called "pure experience." Suppose further that there are, what I shall call "flux-particle laws," that permit the emergence of

quantum properties out of the flux under certain conditions. The Big Bang produced just the right conditions in the flux allowing the quantum particles to emerge. In this case, the quantum particles are properties, ontologically speaking, "over and above" those of the flux. The laws that permit their emergence are themselves basic laws of the cosmos, laws that cannot be explained at the deeper level because there is no deeper level from which they might be explained.

I want to use the concept of 'nomological emergence' to distinguish higher-level properties from lower-level properties and to explain how a lower-level system produces those higher-level properties. In other words, I believe that the occurrence of a higherlevel phenomenon nomologically emerges from a process of interaction of certain lowerlevel properties. I have to be careful not to state that higher-level properties are caused from lower-level properties, because "B is caused by A" and "B nomologically emerges from A" have different meanings. I believe that a clarification is needed especially for the problem of mental causation. One of the features of their differences is that "B is caused by A" is temporal process which means that A occurs prior to B. However, this is not the case for "B nomologically emerges from A." When a person smells food, he feels hungry. In this example, olfactory transduction occurs when odorant molecules reach the olfactory mucosa and bind to the olfactory receptor proteins on the cilia of the olfactory receptor neurons. The receptor protein changes shape which in turn triggers the flow of ions across the receptor-cell membrane and an electrical response is triggered in the cilium. And then, electrical responses in the cilia spread to the rest of the receptor cells, and from there are passed onto the olfactory bulb of the brain in the olfactory nerve. The whole stimulus is a lower-level process and follows a specific physiological law until this

process "causes" the person to *feel* hungry. The feeling of hunger is a nomologically emergent property and is a result of the process of smell, but, strictly speaking, it is not caused by any physiological process. Furthermore, I also believe that the concept of nomological emergence does not violate either the Principle of the Causal Closure of the Physical Domain or the Principle of Causal Exclusion. I will discuss this claim more precisely in 5.5.

The last concept associated with my new concept of emergence is non-structurality. Non-structurality works tightly with causal emergence and novelty. Non-structural emergence stands in contrast to 'structural emergence.' A non-structural emergent property is not merely an aggregation/composition of lower-level entities. It is a genuinely new property produced by that composition, but not one that can be either reduced to or reductively explained by that composition. In other words, an emergent property is not merely a resultant property. Consequently, when non-structural emergent properties have causal effects, those effects are not due to the fact that some mental properties are realized by physical properties. Mental causation is not just physical causation. Neither mental properties nor their causal power are predictable and explainable from the physical level.

To sum up, I would like to give a new definition of emergent property based on the concepts of novelty, nomological emergence, and non-structurality.

- (1) An emergent property is not equal to the totality of its underlying structure.
- (2) An emergent property is distinct from any property of its underlying structure.
- (3) An emergent property lawfully linked to processes of the lower-level configuration of physiological properties when that configuration occurs in an

appropriate circumstance, but the appearance of an emergent property cannot be reductively explained in terms of the causal/functional relationships among the underlying properties.

The first claim and the second claim both are about the concepts of novelty and non-structurality. That last claim expresses the nature of emergence. I think the last claim is compatible with the supervenience thesis too, because say that a higher-level nomologically emerges from a lower-level configuration is to say that there is no higher-level emergence without some change or process at the lower level. Finally, in order for a mental property to emerge, the underlying situation must include not only a physiological process but also the circumstances external to the organism. In other words, mental properties are external properties: if a physical system is not in an appropriate circumstance, the mental phenomenon will not emerge. A defense of this claim is beyond the scope of this dissertation, but I think that this way of conceiving emergent mental properties, makes their postulation more plausible than it might otherwise be.

5.5 Defining Mental Causation within New Emergent Framework

According to the last part of my new concept of emergence, an emergent property is a lawful result of the process of a lower-level configuration within a proper circumstance. The term "result" has a broader meaning of "effect." Here I would like to borrow Ned Hall's two concepts of causation to explain how a mental property is a result of a lower-level configuration. Hall says that we should understand the distinction between "production" and "dependence." He says,

One of these, which I call "dependence", is simply that: counterfactual dependence between wholly distinct events. In this sense, event c is a cause of (distinct) event e just in case e depends on c; that is, just in case, had c not occurred, e would not have occurred. The second variety is rather more difficult to characterize, but we evoke it when we say of an event c that it helps to *generate* or *bring about* or *produce* another event e, and for that reason I call it "production".

Production is better to be used to describe what we ordinarily take to be the relation of causes and effects. Production includes the idea of counterfactual dependence, but it requires something more. At a minimum for causal production, causes and effects must occur at different times. Event B produces event A only if A and B are not simultaneous events, and if A had not occurred, B would not have occurred. For example, to say that my alarm clock going off causes my being late for my flight is to say that if my alarm clock had not gone off, then I would not have been late for my flight, and there is no other reason to cause me being late for the flight within this circumstance. Thus, we say that the alarm clock going off causally produced my being late for the flight in virtue of the fact that my being late for the flight was counterfactually dependent on the alarm clock going off and the two events were distinct in time. I think that causal emergence should be understood as a type of non-productive causation, or as a type of non-causal counterfactual dependence. Exactly which view one takes depends on the view of causation on adopts. A mental property depends on its corresponding physical configuration if and only if they are not overlapping events, and if the corresponding physical configuration had not occurred, this mental property would not have occurred either. Thus, a mental property results from its corresponding physical configuration

_

⁷³Collins, John, Hall, Ned, and L. A. Paul, Causation and Counterfactuals. (The MIT Press, 2004), 225.

under an appropriate circumstance. Therefore, a physical configuration and a circumstance together non-productively cause a mental property to emerge.

Applying my concept of emergence to mental causation, I believe that this new concept is consistent with the doctrines of non-reductive physicalism and also the Principle of the Causal Closure of the Physical Domain and the Principle of Causal Exclusion. My new concept of emergence does not exclude mental causes. According to the Principle of Causal Exclusion, no single effect can have more than one sufficient cause occurring at a given time. Kim uses this principle to exclude mental causes because he believes that physical causes are sufficient to cause every effect. However, based on the concept of nomological emergence, each emergent property nomologically depends on physical properties. Mental causes have to emerge from physical properties to show their causal power. Mental causes alone are not sufficient causes, and for explaining mental causation, physical causes alone are not sufficient causes either. A physical cause and a mental cause are each necessary conditions of an effect and together they are a sufficient cause of the effect. In other words, a sufficient cause is a single event that consists of a single occurrence with both physical properties and emergent mental properties. The accurate causation is depicted at Figure 1.

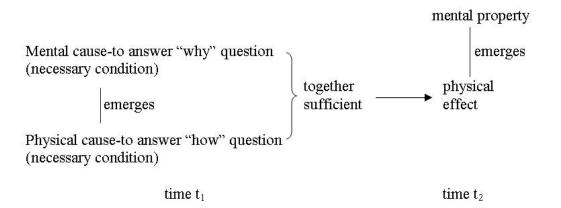


Figure 1: New emergentist theory of mental causation

There is something more than just physical causation. A physical process has to be present in order for mental properties to emerge but neither of them alone is a sufficient to cause some effect.

In short, what I have shown is that mental properties do not need to be realized by physical properties in order to have causal power. To assume that they must assumes in principle the completeness of physics. I do not think that such assumption is fair to mental phenomena. Emergence entails inextricable interactions of mental phenomena in a more complex set of causal processes between and within two systems. In general, if mental causation is a matter of outside physical system, then we need to broaden our eyes to see a different side of this world. Mental properties are real and they have irreducible causal power.

In this dissertation, we have examined in details contemporary substance dualism, the mind-brain identity theory, and Jaegwon Kim's functionalism, and concluded that none of them can provide an appropriate account of mental causation.

By distinguishing the mind from the body, substance dualists face the pairing problem: How does this particular mind unite with this particular body and thus interact with that body? Neither a causal chain nor spatial relationships can connect a mental cause to a physical effect, and there is no other nomological relation if a mind and a body are independent substances. Although three contemporary substance dualists argue that there is a natural bond of a particular mind and a particular body, the same problem remains. Why this body? Why this mind?

The concept of mind as an independent substance gives rise to too many difficulties without compensating explanation. More and more philosophers accept physicalism. The mind-brain identity theorists argue that mental properties are identical with brain properties based on an appeal to ontological simplicity. It entails that mental causation is nothing more than physical causation. Kim argues that the identity relation does not plan any explanatory role because it simply plays a "rewrite" role. The identity relation itself contributes nothing to the explanation of mental causation. It is surprising that the problem of mental causation arises again within the heart of physicalism. It means that accepting a physicalist ontology does not make this problem go away. On the contrary, basic physical assumptions can even be seen as the source of the current difficulties with mental causation.

Kim's functionalism provides functional definitions to link mental properties and their physical realizers. In the end, I argue that Kim still uses physical causal laws to explain mental causation. Every occurrence of mental causation is instantiated at the physical level. Mental properties do not have causal power in their own right.

My preferred idea is that mental properties emerge from physical properties, and both of them together are sufficient to bring about certain effects. The concept of mental emergence is consistent with current science because it does not violate the Principle of the Causal Closure of the Physical Domain or the Principle of the Causal Exclusion. I believe that emergence provides a link to close the gap between the mental and the physical and that is also provides a solution to the problem of mental causation.

Emergence makes mental causation autonomous and also avoids epiphenomenalism

REFERENCES

- Alexander, Samuel. Space, Time, and Deity. Vol. 2, Macmillan and Co, 1920.
- Armstrong, David Malet. "The Nature of Mind." In *Readings in Philosophy of Psychology*, Vol. 1. ed. by Ned Block. Harvard University Press, 1980.
- Armstrong, David Malet. "The Causal Theory of Mind." In *The Nature of Mind*. ed. by David Rosenthal. Oxford University Press, 1991.
- Block, Ned. "What is Functionalism." In *Readings in Philosophy of Psychology*, Vol. 1, ed. by Ned Block. Harvard University Press, 1980.
- Block, Ned. "Troubles with Functionalism." In *Readings in Philosophy & Cognitive Science*. ed. by Alvin Goldman. The MIT Press, 1993.
- Block, Ned. "Anti-reductionism Slaps Back." In *Mind, Causation, World. Supplement: Philosophical Perspectives* 11 (1997): 107-32.
- Block, Ned. and Robert Stalnaker. "Conceptual Analysis, Dualism, and the Explanatory Gap." In *Philosophical Review* 108 (1999): 1-46.
- Block, Ned. "Do Causal Powers Drain Away." In *Philosophy and Phenomenological Research* 67 (2003): 133-50.
- Burge, Tyler. "Mind-Body Causation and Explanatory Practice." In *Mental Causation*. ed. by John Heil and Alfred Mele. Oxford University Press, 1993.
- Chalmers, David. The Conscious Mind. Oxford University Press, 1996.
- Chalmers, David. and Frank Jackson. "Conceptual Analysis and Reductive Explanation." In *The Philosophical Review*, Vol. 110, No. 3 (2001): 315-60.
- Chalmers, David. *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press, 2002.
- Chisholm, Roderick. "Is There a Mind-Body Problem?" In *Philosophical Exchange* 2 (1978): 25-34.
- Collins, John, Hall, Ned, and L. A. Paul. *Causation and Counterfactuals*. The MIT Press, 2004.
- Crane, Tim. "The Significance of Emergence." In *Physicalism and its Discontents*. ed. by Barry Loewer and Carl Gillett. Cambridge University Press, 2001.
- Davidson, Donald. "Actions, Reasons, and Causes." In *The Journal of Philosophy*, Vol. 60, No. 23, (1963): 685-700.
- Davidson, Donald. "Mental Events." In *Mental Causation & the Metaphysics of Mind.* ed. by Neil Campbell. Canada: Broadview Press, 2003a.
- Davidson, Donald. "Thinkging Causes." In *Mental Causation & the Metaphysics of Mind.* ed. by Neil Campbell. Canada: Broadview Press 2003b.
- Fodor, Jerry. "Special Science." In Synthese, Vol. 28, (1974): 97-115.

- Fodor, Jerry. "A Theory of Content and Other Essays. The MIT Press, 1990.
- Foster, John. The Immaterial Self. Routledge, 1991.
- Gillet, Carl. "Does the Argument from Realization Generalize? ---Responses to Kim." In *Southern Journal to Philosophy* 39, (2001): 79-98.
- Harman, Gilbert. "The Inference to the Best Explanation." In *Philosophical Review* 74, (1965): 88-95.
- Hart, W. D. The Engines of the Soul. Cambridge University Press, 1988.
- Hill, Christopher. "Imaginability, Conceivability, Possibility and the Mind-Body Problem." In *Philosophical Studies* 87, (1997): 61-85.
- Hill, Christopher, and Brian McLaughlin "There Are Fewer Things in Reality Than Are Dreamt of in Chalmers's Philosophy." In *Philosophy and Phenomenological Research*, Vol. 59, No. 2, (1999): 445-54.
- Horgan, Terence. "Kim on Mental Causation and Causal Exclusion." In *Nous, supplementary volume, (Philosophical Perspectives, Vol. 11: Mind, Causation, and World)*, (1997): 165-84.
- James, William. "Does 'Consciousness' Exist?" In *The Journal of Philosophy*, Vol. 1, No. 18, (1904): 477-91.
- James, William. The Principles of Psychology. Harvard University Press, 1983
- Jackson, Frank. "Epiphenomenal Qualia." In *Philosophical of Mind: Classical and Contemporary Readings*. ed. by David Chalmers. Oxford University Press, 2002. Originally published in *Philosophical Quarterly* 32, (1982): 127-136.
- Kim, Jaegwon. "'Downward Causation' in Emergentism and Nonreductive Physicalism." In *Emergence or Reduction*. ed. by Ansgar Beckermann, Hans Flohr, and Jaegwon Kim. Walter de Gruyter, 1992.
- Kim, Jaegwon. Supervenience and Mind. Cambridge University Press, 1993a.
- Kim, Jaegwon. "The Nonreductivist's Trouble with Mental Causation." In *Mental Causation*. ed. by John Heil and Alfred Mele. Oxford University Press, 1993b.
- Kim, Jaegwon. "Making Sense of Emergence." In *Philosophical Studies* 95, (1999): 3-36.
- Kim, Jaegwon. Mind in a Physical World. The MIT Press, 2000.
- Kim, Jaegwon. "Multiple Realization and the Metaphysics of Reduction." In *Philosophy of Mind: Classical and Contemporary Readings*. ed. by David Chalmers. Oxford University Press, 2002.
- Kim, Jaegwon. "Blocking Causal Drainage and other Maintain Chores with Mental Causation." In *Philosophical and Phenomenological Research* 67, (2003): 151-76.
- Kim, Jaegwon. *Physicalism, or Something Near Enough*. Princeton University Press, 2005.

- Kim, Jaegwon. *Philosophy of Mind*. Westview Press, 2006.
- Kim, Jaegwon. "Causation and Mental Causation." In *Contemporary Debates in Philosophy of Mind*. ed. by Brian McLaughlin and Jonathan Cohen. Blackwell Publishing Limited, 2007.
- Kim, Jaegwon. "Reduction and Reductive Explanation: Is One Possible Without the Other?" In *Being Reduced: New Essays on Reduction, Explanation, and Causation*. ed. by Jacob Hohwy and Jesper Kallestrup. Oxford University Press, 2008.
- Lewis, David. "An Argument for the Identity Theory." In *The Journal of Philosophy*, Vol. 63, No. 1, (1966): 17-25.
- Lewis, David. "Psychophysical and Theoretical Identifications." In *The Nature of Mind*, ed. by David Rosenthal, Oxford University Press, 1991a. Originally printed in *Readings in Philosophy of Psychology*, Vol. 1. ed. by Ned Block. (Harvard University Press, 1983).
- Lewis, David. "Mad Pain and Martian Pain." In *The Nature of Mind*. ed. by David Rosenthal, Oxford University Press, 1991b. Originally printed in *Readings in Philosophy of Psychology*, Vol. 1. ed. by Ned Block. (Harvard University Press, 1983).
- Levine, Joseph. Purple Haze. Oxford University Press, 2001.
- Loewer, Barry. "Mental Causation, Or Something Near Enough." In *Contemporary Debates in Philosophy of Mind*. ed. by Brian McLaughlin and Jonathan Cohen. Blackwell Publishing Limited, 2007.
- Lowe, E. J. "Causal Closure Principles and Emergentism." In *Philosophy* 75, (2000): 571-86.
- Marcus, Eric. "Mental Causation in a Physical World." In *Philosophical Studies* 122, (2005):27-50.
- McLaughlin, Brian. "In Defense of New Wave Materialism: A Response to Horgan and Tienson." In *Physicalism and Its Discontent*. ed. by Carl Gillett and Barry Loewer, Cambridge University Press, 2001.
- Murphy, Nancey. "Emergence and Mental Causation." In *The Re-Emergence of Emergence*. ed. by Philip Clayton and Paul Davis. Oxford University Press, 2006.
- Nagel, Ernest. "The Reduction of Theories." In *Structure of Science*. Harcourt, Brace & World, Inc. 1961.
- Nagel, Thomas. "What Is It Like to Be a Bat?" In *The Philosophical Review* Vol. 83, No. 4, (1974): 435-50.
- Nagel, Thomas. "Armstrong on the Mind." In *Readings in Philosophy of Psychology*, Vol. 1. ed. by Ned Block. Harvard University Press, 1980.
- O'Connor, Timothy. "Emergent Properties." In *American Philosophical Quarterly* 31, (1994): 91-104.
- O'Connor, Timothy. "Causality, Mind, and Free Will." In Philosophical Perspectives 14,

- Action and Freedom. (2000a): 105-117.
- O'Connor, Timothy. *Persons and Causes: The Metaphysics of Free Will*, Oxford University Press, 2000b.
- O'Connor, Tomothy, and Wong Yu Hong "The Metaphysics of Emergence." In *Nous* 39:4, (2005): 658-78.
- Payment, Paul. "Kim on Closure, Exclusion, and Nonreductive Physicalism." In *Physicalism and Mental Causation*. ed. by Heinz-Dieter Heckmann Fre and Sven Walter. Imprint Academic, 2003.
- Place, U. T. "Is Consciousness a Brain Process?" In *British Journal of Psychology*, Vol. 47, (1956): 44-50.
- Putnam, Hilary. "The Nature of Mental States." In *Philosophy of Mind: Classical and Contemporary Readings*. ed. by David Chalmers. Oxford University Press, 2002. Originally printed in *Mind, Language and Reality: Philosophical Papers*, Vol. 2. ed. by Hilary Putnam. Cambridge University Press, 1975.
- Silberstein, Michael, and John. McGeever. "The Search for Ontological Emergence." In *The Philosophical Quarterly*, Vol. 49, No. 195, (1999): 182-200.
- Silberstein, Michael. "Converging on Emergence: Consciousness, Causation and Explanation." In *Journal of Consciousness Studies*, 8, No. 9-10, (2001): 61-98.
- Silberstein, Michael. "In Defence of Ontological Emergence and Mental Causation" In *The Re-Emergence of Emergence*, ed. by Philip Clayton and Paul Davies, Oxford University Press, 2006.
- Smart, J. J. C. "Sensations and Brain Processes." In *The Philosophical Review*, Vol. 68, No. 2, (1959): 141-56.
- Tye, Michael. "On the Possibility of Disembodied Existence." In *Australasian Journal of Philosophy*, Vol. 61, No. 3, (1983): 275-82.
- Tye, Michael. Consciousness Revisited. The MIT Press, 2009.
- Unger, Peter. All the Power in the World. Oxford University Press, 2006.

VITA

Graduate School Southern Illinois University

Chia-Lin Tu Date of Birth: August 29, 1974

820 W. Freeman St. Apt. 208, Carbondale, IL 62901

1 floor, No. 12-1, Fu-Quei 1st Rd., Pei-tou District 11279, Taipei, Taiwan

chialintu@gmail.com

Soochow University, Taiwan, Bachelor of Art, Philosophy, June 1996

National Chung-Cheng University, Taiwan, Master of Art, Philosophy, June 1999

University at Albany–SUNY, Albany NY, Master of Art, Philosophy, May 2005

Dissertation Title:

The Tale of Mental Causation: Fact or Fiction?

Major Professor: Pat Manfredi

Publications:

(forthcoming) "Is Kim's Approach to Physicalism "Near Enough?"" will be published in *International Journal of Philosophy*, Taipei, Taiwan.

"Is Mental Causation Context-Dependent?" published in *Soochow Journal of Philosophical Studies*, No. 9, March 2004.