

MENTAL CAUSATION, INTENTIONAL ACTION AND EXPLANATORY  
PRACTICE

BY

C2008  
HYUN CHUL KIM

Submitted to the graduate degree program in  
Philosophy and the Graduate Faculty of  
the University of Kansas  
in partial fulfillment of the requirements for  
the degree of Doctor of Philosophy

\_\_\_\_\_  
John Bricke, Chairperson

\_\_\_\_\_  
A. C. Genova

\_\_\_\_\_  
Ben Eggleston

\_\_\_\_\_  
Ann Cudd

\_\_\_\_\_  
Allan Hanson

Date Defended \_\_\_\_\_

The Dissertation Committee for Hyun Chul Kim certifies  
that this is the approved version of the following  
dissertation:

MENTAL CAUSATION, INTENTIONAL ACTION AND EXPLANATORY  
PRACTICE

Committee: \_\_\_\_\_  
John Bricke, Chairperson

Date approved: \_\_\_\_\_

## Abstract

The problem of mental causation results from some unwarranted metaphysical assumption: the Principle of Nomological Character of Causality (NCC). However, there is little reason to understand causation in the manner required to make NCC work. The motivation for the demand for laws in action explanations stems at least in part from the fact that the laws cited in explanations are the laws that subsume events in naturalistic causal relations. By rejecting the idea that causal explanation is causal because it is grounded in natural causal relations, the motivation for requiring laws in explanations disappears. I claim that this is the reason why we need to pay attention to our practice and explanatory strategies. By rejecting NCC we can in fact arrive at a sustainable, defensible and rewarding account of mental causation. The primacy of explanatory practice over the ontological commitment reverses such that an explanation is causal if we accept it as such. By reinterpreting the notion of causation we regain the causal efficacy of the mental.

We look to a theory of intentional action for help in answering the problem of mental causation. In this work I provide a novel conception of intentional action by distinguishing normative reasons from motivating reasons. The proposal recommends itself as being capable of dealing with many problems, including the problems raised by unintended side effects and lucky actions. More importantly, the proposal is able to deal with the problem of causal deviance and consequently is promising in that it avoids epiphenomenalism of mental properties. I conclude the criteria for intentional action must be wide enough to include the normative perspectives of a third-point of view as well as the psychological perspectives.

## **Acknowledgements**

I am deeply grateful to Prof. John Bricke for all the help he has given me in writing this dissertation. He provided extremely valuable comments, often in the form of criticism, which have made this text substantially better than it would otherwise have been. Whatever philosophy I know I have learned mostly from him.

I would also like to thank Prof. Sarah Sawyer for years of stimulating philosophical conversation. I am so grateful to Kevin Dyck, whose friendship sustained me personally throughout the life in the US. I would like to thank my parents who helped me to support in pursuing philosophy.

Finally, but most importantly, I wish to express my deepest, heartfelt appreciation to my family, Doh Yun, Sharon, and especially my loving wife, Jinmi Seo. Throughout this journey, Jinmi has been a faithful companion, a constant encouragement, and a patient support.

I DEDICATE THIS WORK TO GOD.

## TABLE OF CONTENTS

INTRODUCTION	1
CHAPTER	
1. ANOMALOUS MONISM AND THE THREAT OF EPIPHENOMENALISM	7
1.1. ARGUMENT FOR ANOMALOUS MONISM	10
1.2. THE THREAT OF EPIPHENOMENALISM	15
1.3. DAVIDSON'S RESPONSE	20
1.4. OBJECTIONS TO DAVIDSON'S RESPONSE	25
1.5. CONCLUSION	29
2. THE PROBLEM OF MENTAL CAUSATION ARISING FROM CONTENT EXTERNALISM	33
2.1. BURGE'S EXTERNALISM	37
2.2. DAVIDSON'S EXTERNALISM	50
2.3. EXTERNALISM AND THE TOKEN IDENTITY: BURGE AND DAVIDSON	55
2.4. CAUSAL EFFICACY OF THE EXTERNALLY INDIVIDUATED MENTAL CONTENT: BURGE VS. FODOR	68
2.5. CONCLUSION: REJECTING NCC	73
3. THE EXCLUSION ARGUMENT AND THE PRINCIPLE OF THE NOMOLOGICAL CHARACTER OF CAUSALITY	78
3.1. SOME INTUITIONS AGAINST NCC	81
3.2. THE EXCLUSION ARGUMENT	84
3.3. AGAINST THE EXCLUSION ARGUMENT	92
3.4. AN ARGUMENT AGAINST CIP: CONSIDERATION FROM CONTENT EXTERNALISM	97
3.5. THE EXCLUSION ARGUMENT: CONSIDERATIONS FROM EXPLANATORY PRACTICE	102
4. COMMONSENSE PSYCHOLOGY AS AN EXPLANATORY PRACTICE	107
4.1. COMMONSENSE PSYCHOLOGY (CP) NOT AS A SCIENTIFIC THEORY BUT AS A PRACTICE	111
4.2. RESPONSE TO THE FIRST CHALLENGE	118
4.3. RESPONSE TO THE SECOND CHALLENGE	121
4.4. RESPONSE TO THE THIRD CHALLENGE	127

4.5. CONCLUSION	140
5. INTENTIONAL ACTION AND NORMATIVE CONSIDERATIONS	142
5.1. THE STANDARD ACCOUNT OF INTENTIONAL ACTIONS AND ITS DIFFICULTIES	145
5.2. THE CHALLENGE TO THE STANDARD ACCOUNT: SKILL/LUCK	150
5.3. THE CHALLENGE TO THE STANDARD ACCOUNT: UNINTENDED SIDE EFFECTS	153
5.4. NORMATIVE/MOTIVATING REASONS	156
5.5. THEORETICAL GROUND FOR NORMATIVE CONSIDERATIONS	159
5.6. CONCLUSION	164
6. NORMATIVE CONSIDERATIONS AND MENTAL CAUSATION	167
6.1. INTENTIONAL ACTION	170
6.2. UNINTENDED SIDE EFFECTS REVISITED	175
6.3. SKILL/LUCK REVISITED	179
6.4. CAUSAL DEVIANCE	185
6.5. NCC AND MENTAL CAUSATION DEBATE REVISITED	195
6.6. CONCLUSION	201
BIBLIOGRAPHY	205

## INTRODUCTION

The concept of intentional action is connected with that of reasons. Some philosophers define a purposeful, intentional action as one which is done for a reason. But the problem is that there are intentional actions that are not done for a reason and there are actions done for reasons that still are not intentional. In this work I provide a novel conception of intentional action by distinguishing normative reasons from motivating reasons. The conception is as follows:

[Intentional action] An agent's  $\Phi$ -ing is intentional iff either (i) it is done for her motivating reason (if it is not the case of luck or causal deviance) or (ii) the fact that certain consequences would occur was a justifying reason not to perform the action.

The definition should be reflected on both reasons. The proposal recommends itself as being capable of dealing with many problems, including the problems raised by unintended side effects and lucky actions. More importantly, the proposal is able to deal with the

problem of causal deviance and consequently is promising in that it avoids epiphenomenalism of mental properties.

The problem of mental causation emerges when we want to confer some kind of primacy to the physical without abandoning the autonomy of the mental. The nonreductive physicalist who holds that the mental is causally efficacious needs to show how it is that mental properties themselves can make a causal difference without at the same time rendering themselves reducible to physical properties.

Chapter One discusses a problem of mental causation by exploring Donald Davidson's Anomalous Monism (AM). I show that Davidson runs into difficulties when it comes to accommodating our commonsense intuitions about the nature of mental causation. So long as Davidson holds the Principle of Nomological Character of Causality (NCC), I argue, he is left with the following dilemma: either he treats the mental as causally efficacious and therefore gives up our commitment to the idea that the mental realm is irreducible, *sui generis*, or he holds onto that latter notion, but jettisons the intuition that our mental states are causally efficacious. (Either Reduction or Epiphenomenalism.) I claim that we should accept both the



intuition that the mental is anomalous and that it is causally efficacious. I will claim that NCC is not something that we can tolerate.

In Chapter Two I will deal with a tension that arises from content externalism. This is the problem, resulting from the seeming conflict between the two claims, one that ordinary psychological states play causal roles in psychology in virtue of their contents, and the other that their contents are, in part, individuated by the nature of their referents. I will examine a debate between Davidson and Burge. Considering that debate both will strengthen my claim in Chapter One, that AM is committed to the epiphenomenalism of the mental, and therefore that NCC should be rejected, and will help to elucidate content externalism in general. By examining a debate between Burge and Fodor, I argue that that there is no *a priori* reason why the so-called "wide" contents do not or cannot play causal roles in psychological explanations of behavior, and show how they might do so by noting that wide contents are among the properties we ordinarily cite to explain our behavior.

The result we elicit from both debates, one between Burge and Fodor (the issue of the compatibility of

externalism with the causal efficacy of the mental) and the other between Burge and Davidson (the issue of the compatibility of externalism with the token identity thesis) is that we have good reasons for rejecting NCC.

A general solution to the problem of mental causation arising from content externalism can enable us to see how such a solution helps to solve the problem of the Exclusion Argument, which is the subject of Chapter Three. The Exclusion Argument is designed to show that nonreductive conceptions of the mental face the serious problem of producing an account of mental causation which does not render the mental epiphenomenal. I argue that the Exclusion Argument is not successful. The rejection of the argument is reached by the rejection of the Causal Inheritance Principle (CIP), which says that a mental property, realized in virtue of a physical realization base, has no new causal powers beyond the causal powers of its physical base. This is important because the rejection of CIP entails the rejection of NCC.

In the previous chapters I argued that a particular unanalyzed assumption, NCC, is responsible for a philosophical impasse. In Chapter Four, I will describe the new conception of causation that emerges as a result

of rejecting NCC. In this chapter I will first explain commonsense psychology (CP), and then argue against the claim that CP is a kind of a scientific theory. The alternative to regarding CP as a scientific theory is to regard it as a practice. Secondly, I will argue that our explanatory practice should guide our ontological commitment. And, finally, I will defend my position against what I see to be a number of serious challenges.

The new conception of causation that emerges as a result is strengthened by a theory of intentional action that I will endorse in the final two chapters. In Chapter Five I will provide a theoretical ground to include normative perspectives in dealing with the concept of intentional actions. I will claim that our ordinary practice in attributing intentional action in particular cases, and our practice of attributing reason explanations, can actually be influenced by normative considerations. I set the stage by examining some of the problems associated with the concepts of intentional action that are frequently discussed in the literature in the philosophy of action. I will provide an explanation of understanding intentional action by invoking the concepts of motivating reason and justifying reason.

In the final chapter I provide a necessary and sufficient condition for intentional action by developing the idea of the previous chapter. The definition pays close attention to normative considerations as well as motivating reasons. The definition proves itself capable of solving a number of other problems related to intentional actions, including the problems of unintended side effects, deviant causal chains, and skill. Most importantly, it provides a way of understanding the problem of mental causation. Because normative considerations play a role in determining whether an action was performed intentionally, I claim that it is difficult to see how NCC can be true.

**CHAPTER 1**  
**ANOMALOUS MONISM AND THE THREAT OF**  
**EPIPHENOMENALISM**

The traditional problem of mental causation, the so-called Cartesian problem, is a conflict between the intuition that the mind and the body are radically different things and the intuition that the mind and the body causally interact. If the mind and the body are two distinct kinds of substances that can exist independently of each other, it is hard to explain how the mind and the body interact causally.

The contemporary problem of mental causation, though different from the Cartesian one, emerges from related intuitions. It is different because the nature of the mental and its relation to our bodies is discussed nowadays in terms of mental properties of physical organisms. However, the problem of mental causation is not abolished by eliminating substances; it reappears when we want to confer some kind of primacy to the physical without abandoning the autonomy of the mental. We could, some would, claim that we have to get rid of the mental or to identify it with the physical. In this

case we don't need a further account of the mental and, in particular, we don't need to deal with the issue of how the mental causally interacts with the physical. Nevertheless, as it happens, a vast majority of contemporary views want it both ways: the physical is primary but the mental is real and distinct from it. And this is the arena in which problems similar to the Cartesian one emerge.

This chapter discusses a problem of mental causation. With Donald Davidson's well known theory of the mind, Anomalous Monism (hereafter AM) as my concrete example of nonreductive physicalism, I shall devote the remainder of the chapter to showing that nonreductive physicalism runs into difficulties when it comes to accommodating our common sense intuitions about the nature of mental causation.

I use Donald Davidson's anomalous monism as an example of nonreductive physicalism for a couple of reasons. First of all, Davidson is the philosopher who has made famous both the idea that mental and physical vocabulary operate with different constitutive standards, and the idea that the best way to make sense of the idea that one's beliefs and desires can explain one's behavior

is to recognize that they caused that behavior. Second, and more important, Davidson's theory stands as one of the most worked out attempts to accommodate both of these ideas in one comprehensive account of the mental.

Therefore, this chapter is devoted to an explication of Donald Davidson's AM in detail and discusses its problems with regard to the charge of epiphenomenalism. Many critics argue that AM does not save causal efficacy for mental events as mental. In the subsequent sections of the chapter I will present Davidson's responses to the objections that his view makes the mental causally inefficacious.

In his 1993 paper "Thinking Causes," Davidson, for the first time, addresses the worries expressed by Kim and others. In so doing, Davidson claims that his critics' talk of mental properties making or not making a causal difference is at odds with the extensionalist conception of causal relations that he advocates. Given the clearly Quinean ontological framework within which he works, Davidson does not admit properties into his ontology, and, therefore, claims that the objections rest on a fundamental misunderstanding of some aspect of his view. Second, Davidson finally explains how, according to

AM, the mental can in fact be causally relevant<sup>1</sup>. I will discuss Davidson's point and claim that Davidson's explanation is unsuccessful. I will argue that the epiphenomenalist objection succeeds in identifying a serious problem for AM.<sup>2</sup>

### 1.1 ARGUMENT FOR ANOMALOUS MONISM

In this section I will focus on Donald Davidson's AM as presented in a series of influential articles reprinted in his *Essays on Actions and Events*. I will present Davidson's argument for AM as the identity theory which is entailed by the consistency of the three principles. I will explain how Davidson thinks the three principles are to be reconciled and why he thinks they

---

<sup>1</sup> I will call a thing causally efficacious if it is able to cause another thing to occur, and causally relevant if it is able to explain why something has happened due to some cause. For example, if *c* causes *e*, then we say *c* is causally efficacious in regard to *e*; if *c* can explain why *f* has occurred or what has caused *f* to occur, then we say *c* is causally relevant in regard to (the causing of) *f*. But we cannot say *c* is causally relevant *simpliciter*. The expression "in regard to (the causing of) *f*," is indispensable with causal relevance. In light of this usage, causal efficacy is a metaphysical or ontological notion while causal relevance is an explanatory one. This usage implies that causal efficacy and causal relevance are different in at least one significant sense: causal efficacy may ground a causal relation and causal relevance is grounded by a causal relation.

<sup>2</sup> In Chapter Three I shall deal with the so-called Exclusion Argument. I contend, following critics, that AM succumbs to the Exclusion Argument.



imply the token identity<sup>3</sup> of mental events with physical events.

Davidson finds each of the following principles (Davidson 1970: 208) to be plausible and very likely to be true:

[The Principle of Causal Interaction]: At least some mental events interact causally with physical events.

[The Nomological Character of Causality]: Events related as cause and effect fall under strict, deterministic laws.

[The Anomalism of the Mental]: There can be no strict deterministic laws on the basis of which mental events can be predicted and explained.

If one were to accept all three principles as true, one would be faced with explaining their apparent inconsistency. For it is natural to read the first two principles as entailing the denial of the third. If at least some mental events are related as cause or effect with physical events, and where there is causation there is subsumption by law, then it seems there must be a law which subsumes the mental and physical events.

---

<sup>3</sup> Davidson's version of the identity thesis does not entail that all mental properties are also physical properties; only causal properties of events, however else described, enter the proper domain of physical explanations. MacDonald says a similar point: "[T]he argument works to establish token identity of the mental and the physical only for those mental events which ... interact causally with physical events" (MacDonald 1989: 87).

Since Davidson holds that the three principles are indeed true, their incompatibility must be only apparent. Briefly, the solution to the apparent inconsistency is as follows: Causality and identity are relations that obtain between individual events independently of descriptions. The Principle of Causal Interaction (hereafter CI) applies to events in extension and so is independent of whether they have physicalistic or mentalistic descriptions (Davidson 1970: 215). Thus if  $e$  causes  $f$  then those two events are in that causal relation whether we say so by describing  $e$  as Jack's fall and  $f$  as a disaster or by describing  $e$  and  $f$  using different descriptions. Thus, CI concerns events in extension and "is therefore blind to the mental-physical dichotomy" (Davidson 1970: 215).

The Anomalousness of the Mental (hereafter AME) concerns events described as either mental or physical; it does not concern events *per se*, i.e., individual events or event-tokens. AME ensures the anomalousness of the mental by denying that strict laws under which an event can fall are formulable when that event is described in mental terms. AME, therefore, should be read as saying that there are no strict laws which connect

events under mental descriptions with other events. That is to say, no singular causal statement which refers to an event *via* a mental description instantiates a strict law, and no generalization which makes essential use of mental descriptions to refer to events can ever be a law.

Consider the Principle of the Nomological Character of Causality (hereafter NCC). What might Davidson mean in saying that two events "fall under a strict law"? We can think of falling under a law as the same thing as being "covered" or "subsumed" by a law. But laws, as Davidson points out, are linguistic in that they necessarily refer to events via descriptions. Thus, if laws are linguistic, to say that two causally connected events "fall under" or are subsumed by a law is to say that they have descriptions (whether or not we can pick those descriptions out) such that the singular causal statement connecting them under those descriptions instantiates a law. On this interpretation, then, Davidson's NCC does not imply that every singular causal statement instantiates a law, but is consistent with there being true singular causal statements that do not instantiate any laws.

Now we should be able to see that CI, NCC and AME are consistent with one another. CI and NCC do not entail that there are strict laws which connect mental events under mental descriptions with physical events under physical descriptions, which would be the denial of AME. Rather, together they imply only that when a mental event is causally connected with a physical event, there will be descriptions of those two events such that the singular causal statement connecting those two events under those descriptions instantiates a strict law.

Now we can see that given AME, those descriptions cannot be mental descriptions. It follows, then, that those descriptions must be physical descriptions. Thus, given Davidson's account of what it is for an event to be a mental event or physical event, those events subsumed by strict law are physical events. Thus, we have the token identity of mental events (at least those which causally interact with other events, either mental or physical) with physical events. The view which results from this reconciliation of the three principles is what Davidson calls Anomalous Monism (AM). This is the view that although mental events are physical events, there

are no laws strictly correlating the mental with the physical.

## 1.2 THE THREAT OF EPIPHENOMENALISM

In this section I examine the charge that AM is committed to epiphenomenalism. The charge questions the consistency of CI and the other two principles, NCC and AME. I demonstrate how AM face difficulty in making adequate sense of causal efficacy of the mental. I will present critics' attack offered by Honderich and Kim, and will show that the criticisms do make sense in charging AM with epiphenomenalism. I will then explain why I think AM necessarily renders the mental causally inert. This insight will point us in the direction of a solution to the epiphenomenalist attack.

In recent discussion of AM there has been some question as to whether the view is committed to the epiphenomenalism of the mental. The worry is not that AM renders mental *events* causally inert, for mental events are token-identical with some physical events on Davidson's account; the charge is rather that mental *properties* of mental events have no causal role to play

under AM. It is the physical properties in question that the causal work is being performed.

I examine Honderich's argument in detail. His argument is that when one event causes another event, it makes sense to ask which properties of the two events were relevant to their being in causal relation. That is, it is always acceptable to ask which properties of the former are causally relevant to its being the effect of the latter event. With respect to the relation between the mental and the physical, the question is whether it is the mental or the physical properties of a mental event which are causally relevant. That is, is it the mental as mental or the mental as physical which is causally efficacious? If it is answered by saying that it is the mental as mental which is causally relevant, then AM must reject AME - there must be psychophysical laws. If it is answered by saying that it is the mental as physical which has causal power, then CI comes into question since our initial acceptance of it was based on the natural understanding of it as saying that the mental

as mental causally interacts with the physical, and AM seems to be committed to Type-Epiphenomenalism.<sup>4</sup>

Let me take a look at Honderich's argument in detail. Honderich (1982) argues that the three principles on which AM is grounded are incompatible when it is recognized that there are indefinite numbers of ways to express an event and therefore only certain properties of events are causally relevant to their being the causes or effects that they are. He argues that the recognition of causally relevant properties raises a question of the

---

<sup>4</sup> There are two kinds of epiphenomenalism that mental might be causally inert. According to the first, while certain events have both mental and physical characteristics, those events never cause other events in virtue of having those mental characteristics but only in virtue of having the physical characteristics they do. Brian McLaughlin calls this Type Epiphenomenalism (Type-E) and defines it as follows:

[Type-E] (a) Events can be causes in virtue of their physical properties, but (b) events cannot be causes in virtue of their mental properties. (McLaughlin 1989: 108).

The second kind of epiphenomenalism is the view that no single event has both mental and physical characteristics (i.e., no single event is both a mental and a physical event), and that while every mental event is caused by some physical event no mental event is ever a cause of any other event, either mental or physical. McLaughlin identifies this view as Token Epiphenomenalism (Token-E) and defines it as follows:

[Token-E] (i) Physical events can cause mental events, but (ii) mental events have no causal powers; they cannot cause mental events, nor can they cause physical events. (McLaughlin, 1989: 110).

Davidson is able to deny Token-E. Critics, however, have argued that AM is committed to Type-E.

legitimacy of AM. The unhappy results come when we realize that it does make sense to ask whether it is a mental event as mental that causes a physical event or the mental event as physical causes the event. If the first route is the route that anomalous monists take, then they have the denial of AME and therefore the denial of AM itself. If, on the other hand they take the second route in order to keep AME, then they must give up CI that there is causal connection between the mental as mental and the physical.

Honderich points out that it does make sense to talk of something's being such and such *under a description*. He says, "To talk this way is to speak of certain properties of a thing rather than others. To say two things are not in lawlike connection under certain descriptions is to say that certain of their properties are not in lawlike connection, or, perhaps, that the things are not in lawlike connection in virtue of certain of their properties." (60-61) It is clear that it is certain properties of the event which are relevant to its being the cause it is. Honderich gives an example of moving the scale to the two-pound mark by putting green and French pears on the scale. The event of putting



something that is green and French did cause the event of moving the pointer to the two-pound mark. In this case, however, it does not make sense to say that because of the pears' greenness and Frenchness the pointer moves to the two-pound mark. There is in fact no entailed law connecting the event *in virtue of* its being of something green and French with the pointer's so moving.

In the above example of pears, neither the greenness nor the Frenchness of the pear does not cause the pointer's movement, rather the weight of the pears does cause it. Then, there is no difficulty in saying that it is in virtue of certain of its properties rather than others that an event is the cause it is. The causal connection holds between the weight of the pears and the movement of the scale. Even though the greenness and Frenchness of the pears make the event what it is, those properties are not necessary to the event's being the cause it was. From the above consideration, Honderich elicit the following principles:

[The Nomological Character of Causally Relevant Properties]: It does follow from the fact that  $E_1$  caused  $E_2$  in virtue of a property  $f$  of  $E_1$  and property  $g$  of  $E_2$  that  $E_1$  and  $E_2$  are in lawlike connection partly or wholly in virtue of properties  $f$  and  $g$ . ]

So given that not all properties of an event are relevant to its being the cause or effect of another event, the question arises what properties are relevant. Namely, the question is whether it is the mental or the physical properties of a mental event which are causally relevant. That is, is it the mental as mental or the mental as physical which is causally efficacious? If it is answered by saying that it is the mental as mental which is causally relevant, then AM must reject AME - there must be psychophysical laws. If it is answered by saying that it is the mental as physical which has causal power, then CI comes into question since our initial acceptance of it was based on the natural understanding of it as saying that the mental as mental causally interacts with the physical, and AM seems to be committed to Type-Epiphenomenalism.

### **1.3 DAVIDSON'S RESPONSE**

Davidson Have defended AM by essentially claiming that these criticisms are based on an assumption about the relation among descriptions, events, and causal laws

which Davidson does not and should not accept.

Specifically, it rests on the assumption that events have descriptions and are thereby subsumed by causal laws in virtue of having certain properties. Given Davidson's ontological framework that the relata of causation are events and his concept of events is purely extensional, it would be unfair to attack AM on the grounds that it makes mental properties epiphenomenal: it is unfair to ask whether events are subsumed by causal laws in virtue of their properties because in Davidson's ontology he does not assume the existence of properties, therefore it is events in extension which are in lawlike connection and not events under certain descriptions.

Davidson does argue that he does not accept this assumption and is actually committed to its denial. However, since such an assumption is necessary if the charge of epiphenomenalism is to apply to AM and Davidson does not make that assumption in arguing for AM, he claims that AME and NCC cannot be shown to be inconsistent with CI.

Honderich argued it is always an appropriate question to ask which properties of events are properties in virtue of which they are causally related, and hence

events related as cause and effect are in lawlike connection in virtue of certain of their properties. Can he make this same move without assuming the existence of properties? Davidson thinks not. For Honderich would have to be able to show that it is always a relevant question either to ask which events are events in virtue of which two events are causally related, or to ask which descriptions of events are descriptions in virtue of which they are causally related. But the answer to the first question is trivial, since clearly it is just those two events which are causally related which are relevant to their being so related. And the second question makes no sense, since it is events in extension which are in causal connection and not events under certain descriptions.

This point becomes clear when we examine the debate between Davidson and Kim. If one holds, as Kim suggests Davidson ought to, that NCC entails that it is only in virtue of falling under a physical law that an event causes, then one would seem to be in the position of having to admit that an event's mental properties can't make a causal difference. It would follow then that NCC does imply that the mental is not causally efficacious.

However, if one resists Kim's interpretation of NCC then perhaps the situation won't seem too dire. Or at least, this is what Davidson wants to argue. Why would someone be led to believe, Davidson asks, that NCC entails that it is only in virtue of falling under a physical law that an event can cause? According to Davidson, one can only arrive at such a conclusion if one makes the mistake of reading him as saying that, on AM, "events are causes or effects only as they instantiate physical laws" (Davidson 1993: 13).<sup>5</sup> But, Davidson now reminds us, on his account events are non-abstract particulars, which means that causal relations are extensional. To say that a relationship is extensional is to leave no room for the concept of "cause as," a concept which would make causality an intensional relation. For Davidson causal relation holds between events no matter how they are described:

It is *events* that have the power to change things, not our various ways of describing them. Since the fact that an event is a mental event, i.e. that it can be described in a psychological vocabulary, can make no difference to the causes and effects of that event, it makes no sense to suppose that describing it in the psychological vocabulary might deprive the event of its potency. (Davidson 1993: 12)

---

<sup>5</sup> As we saw, Kim (1989) attributes this position to Davidson.

Redescribing an event therefore cannot, Davidson says, change its causal efficacy:

If causal relations and causal powers inhere in particular events and objects, then the way those events and objects are described, and the properties we happen to employ to pick them out or characterize them, cannot affect what they cause. (Davidson 1993:8)

This means that Kim is wrong to suggest that NCC entails that events cause in virtue of their physical properties, but not in virtue of their mental properties. Strictly speaking, on Davidson's view it is "events that have causes and effects" (Davidson 1993: 13). The fact that events stand in causal relations does not, therefore, depend on any properties, mental or physical, which can be ascribed to them. We are now in a position to see why Davidson claims that Kim's charges rest upon a confusion concerning the nature of causation and causal explanation. For Davidson causation is an extensional relation that holds between events, regardless of how they are described. On the other hand, causal explanation involves describing an event in such a way that it fits into some larger pattern of events; such a pattern might be

physical (nomological) or mental (rational). By conflating causation and causal explanation Kim has imposed an unjustifiable restriction on Davidson's account.

#### 1.4 OBJECTIONS TO DAVIDSON'S RESPONSE

As we saw, Davidson makes the claim that no event can cause anything in virtue of its mental properties or its physical properties. It is not at all clear that such a view is consistent with NCC<sup>6</sup>, which itself seems to implicate the physical properties of an event. However, there are some problems that Davidson's AM faces.

McLaughlin examines Davidson's extensional view of causal relations, according to which it makes no literal sense to speak of causing an event in virtue of their properties. McLaughlin claims that Davidson is mistaken in holding that C1 incompatible with C2:

(C1) The *relata* of the causal relation are non-abstract, particular events; and if event *c* caused event *e*, and *c=d*, then *d* caused *e*; and if *c* caused *e*, then there is something that caused *e*. (1993: 30-31)

---

<sup>6</sup> I will discuss NCC in full in Chapter Three.

(C2) If event *c* caused event *e*, then *c* caused *e* in virtue of certain of *c*'s properties. (1993: 31)

McLaughlin wants to argue first, that C1 and C2 are in fact consistent, and second, that C2 can be literally true. If McLaughlin can support both of these claims then he will have succeeded in demonstrating that Davidson is not justified in claiming that events do not cause in virtue of their properties.

Davidson's own example in "Thinking Causes" to support the claim that C1 and C2 are inconsistent is the extensional relation between non-abstract particulars, the weighs-less-than relation. Davidson would think that the following two claims are inconsistent, but McLaughlin claims that they are not:

(W1) The relata of the weighs-less-than relation are non-abstract, particular substances; and if *a* weighs less than *b*, and  $a=c$ , then *c* weighs less than *b*; and if *a* weighs less than *b*, then there is something that weighs less than *b*. (1993: 31)

(W2) If substance *a* weighs less than substance *b*, then *a* weighs less than *b* in virtue of certain of *a*'s properties. (1993: 32)

McLaughlin think that the two claims in (W1) and (W2) can be consistent, namely the extensional view of weighs-



less-than relation between non-abstract particular substances holds in virtue of certain of something about each, namely, their weights.

This point carries over to the "causes" relation as well: there is no inconsistency in holding that (i) the "causes" relation is an extensional one holding between non-abstract particular events, and (ii) that if one event causes another event, it does so in virtue of certain of its properties (McLaughlin 1993: 31).

Why would Davidson think that C1 and C2 are inconsistent? Davidson seems to argue that if one believes that one event causes another event in virtue of one of its properties, or in virtue of belonging to a certain type commits one to the view that in order to be true a singular causal statement relating those events must describe them in terms of those very same properties or types. But this is not the case. Acceptance of the fact that C1 and C2 are consistent does not commit one to holding that singular causal statements are only true if they themselves specify the relevant causal properties, a view that Davidson clearly cannot allow. It is Davidson's failure to recognize this point, McLaughlin contends, which leads him to argue that C1 and C2 are inconsistent.

Even if we can show that C1 and C2 are consistent, it might still be possible to argue that C2 is nevertheless false. This would salvage Davidson's position, but is this route available to him? One reason we have for thinking that Davidson might want to adopt such a strategy is that, according to McLaughlin, Davidson appears to think that if *c* causes *e* in virtue of *c*'s having *F*, then it would follow that "*c*'s having *F* causes *e* (or that *c* causes *e* under the description 'the *F*')"

(McLaughlin 1993: 33). Such a scenario would indeed be problematic on Davidson's account because "*c*'s having *F*" is a state of affairs rather than an event, which means that the causal relation would no longer be an extensional one.<sup>7</sup> But, according to McLaughlin, such an implication does not follow. Saying that an event causes something in virtue of one of its properties actually implies that the event itself is a cause (McLaughlin

---

<sup>7</sup> Relations between states of affairs are not extensional because the truth-value of such sentences can change depending on how such states of affairs are described. For example, while it may be true that Oedipus's having the attitude of *wanting to marry* Jocasta caused him to marry a particular woman, it would not be correct to say that Oedipus's having the attitude of *wanting to marry his mother* caused him to marry a particular woman, even though in his case the terms 'Jocasta' and 'his mother' are co-referential. Such contexts are referred to as opaque (as opposed to transparent) contexts.

1993: 33). Again McLaughlin uses Davidson's "weighs-less-than" example to demonstrate his point:

That *a* weighs less than *b* in virtue of weighing 10 pounds, does *not* imply that *a*'s weighing less than 10 pounds weighs less than *b*. (McLaughlin 1993: 33-4)

Just as there is no danger in this example that objects, *a*, and states of affairs, *a*'s weighing less than 10 pounds, will be confused with each other, there is no danger, when it comes to causation, that events and states of affairs will get confused with each other. To be more specific:

The claim that event *c* caused event *e* in virtue of *c*'s having *F* does not imply that the state of affairs consisting of *c*'s having *F* caused *e*. (McLaughlin 1993: 34)

So it turns out that *C2* is not false, and that *C1* and *C2* are in fact consistent with each other. Davidson is therefore not justified when he argues that it makes no sense to speak of an event's properties making a causal difference.

## 1.5 CONCLUSION

I believe McLaughlin's claim is persuasive. The Nomological Character of Causality (NCC) leads to epiphenomenalism because it stipulates that the only way something can be causally relevant is for it to be a physical property. Davidson denies this, but only because he mistakenly thinks that an extensionalist view of causation precludes properties from themselves playing a causal role. But McLaughlin shows why Davidson is incorrect. Indeed, this causes so much difficulty for Davidson that he is forced into the counterintuitive position of having to argue that properties can make a difference even though events don't cause in virtue of their properties.

McLaughlin thinks that NCC is the culprit, since it holds that the only way something can be causally relevant is by falling under a physical type. Kim clearly holds on to NCC, or at least to the view that causation always involves the notion of kinds of events being in relation to each other. Thus he claims that questions of the form "What is it about events  $c$  and  $e$  that makes it the case that  $c$  is a cause of  $e$ ?" can be answered by saying that " $c$  is an event of kind  $F$  and  $e$  is one of kind

G (and, you may add if you favour a nomic conception of causality, there is a law of an appropriate form connecting F-events with G-events)" (Kim 1993a: 22). We do so because, according to Kim, we need to acknowledge that "the causal relation obtains between a pair of events because *they are events of certain kinds, or have certain properties*" (Kim 1993a: 22).

It now seems as if we are confronted with the following. The nonreductive physicalist who also holds that the mental is causally efficacious needs to show how it is that mental properties themselves can make a causal difference without at the same time rendering themselves reducible to physical properties. But so long as we hold onto NCC it will appear that this can't be done. So if we hold onto NCC we are left with the following dilemma: either we treat the mental as causally efficacious and therefore give up our commitment to the idea that the mental realm is irreducible, *sui generis*, or we hold onto that latter notion, but jettison the intuition that our mental states are causally efficacious. Since both commitments are powerful ones, we are left with an intolerable situation. Whichever way we lean, it appears that we must sacrifice part of our commonsense conception

of the mental. I claim that we should accept both the intuitions that the mental is anomalous and that it is causally efficacious, but not in the Davidsonian way. I will take up the issue of NCC in detail by dealing with the problem of mental causation generated by the extrinsic nature of mental content and the one generated by the Exclusion Argument. I will claim that NCC is not something that we can tolerate in the course of dealing with the two problems.

**CHAPTER 2**  
**The PROBLEM OF MENTAL CAUSATION ARISING FROM**  
**CONTENT EXTERNALISM**

The central idea of content externalism is that the contents of mental states are not determined exclusively by what occurs in us but are determined in part by external states of affairs. Although there is still a debate whether externalism itself is true, a number of recent investigations have begun to explore the question of what follows if it is true. In this chapter I will deal with a tension that arises from content externalism. This is the problem resulting from the seeming conflict between the two claims, one that ordinary psychological states play causal roles in psychology in virtue of their contents, and the other, *content externalism*, that their contents are, in part, individuated by the nature of their referents: what causes me to drink water, it might be maintained, is some neurophysiological property of me; the fact that I am environmentally related to water and not to T-water bears no lawlike relationship with my action; if content properties enter into no genuine laws governing the causation of action, it may be argued, then

content-based explanations are not causal in nature. I will support the claim that beliefs and other mental states with widely individuated intentional contents play genuine causal roles in virtue of their contents in psychological causal explanations of behavior.

In section 2.1, I discuss Burge's famous *Twin-Earth thought experiment*, the central aim of which is to show that content externalism is a metaphysical view about the nature of certain mental states – what having such states necessarily presupposes. In Chapter One I chose AM as an example of nonreductive physicalism. One of the reasons I chose AM is that it stands as one of the most worked out attempts to accommodate both of the ideas, the physical is primary but the mental is real and distinct from it, in one comprehensive account of the mental. As we saw, however, critics showed that the three principles Davidson used to elicit AM are not consistent; they showed AM to be committed to a version of epiphenomenalism. Those, like me, comfortable with the rejection of AM, however, still want to confer some kind of primacy to the physical without abandoning the autonomy of the mental. Because Davidson's AM is a monism, claiming an identity between mental and physical events,



someone like Burge may view Davidson's position as presupposing something against externalism.<sup>8</sup>

Before dealing with the issue between externalism and nonreductive physicalism, in section 2.2 I will present Davidson's own brand of externalism and his rejection of Burge's Twin-Earth thought experiment in general. In section 2.3 I will examine the debate between Davidson and Burge for the following reasons: (1) the result of the debate will strengthen my claim in Chapter One, that AM is committed to the epiphenomenalism of the mental, and therefore that NCC should be rejected; and (2) the debate helps to elucidate content externalism in general. I present Burge's argument against the token identity thesis (1993; 1979). Burge attacks Davidson by arguing that Davidson cannot consistently hold both AM and content externalism. Davidson attempts to show that this is not the case by introducing his so-called *Sunburn Argument*. In this section I will argue that the Sunburn Argument does not work. As a result of the argument against the token identity thesis, Burge rejects NCC.

---

<sup>8</sup> As I will mention in section 2.2 Davidson's own brand of externalism differs in relevant ways from what has been generally called "externalism," particularly that of Tyler Burge. It is indeed an interesting matter to see whether Davidson's externalism is compatible with content externalism. The issue is complex, and requires more development than I can undertake in this work.

Burge claims that we do not know and cannot know *a priori* that causal statements entail the existence of strict laws. There is no reason to think that unless mental causation is just physical causation it would interfere with physical processes.

Section 2.4 is the main section of this chapter. It contains debates between Burge and Fodor. By examining the debates I argue that there is no *a priori* reason why so-called "wide" contents do not or cannot play causal roles in psychological explanations of behavior, and show how they might do so by noting that wide contents are among the properties we ordinarily cite to explain our behavior.

The final section, section 2.5, presents one interesting result I elicit from both debates, one between Burge and Fodor and the other between Burge and Davidson. It is the rejection of the Principle of the Nomological Character of Causality, one of three premises Davidson takes to be true to argue for AM. In fact if there is no good reason to accept NCC, it follows, I argue, that content externalism is compatible with the causal efficacy of the mental. In Chapter Three I will show that the solution to the Exclusion Argument can be

reached by rejecting the Causal Inheritance Principle (CIP). We will see that the rejection of CIP actually implies that NCC is in fact wrong.

## 2.1 BURGE'S EXTERNALISM

Burge's thought experiment is designed to show that so-called "anti-individualism"<sup>9</sup> is a metaphysical view about the nature of certain mental states – what having such states necessarily presupposes. Burge's conclusion rests on a three-step thought experiment. In this section I will deal with each of these steps in detail in order to better understand two issues, implicit in externalism: the issue of the compatibility of externalism with the token identity thesis (section 2.3); and the compatibility of externalism with the causal efficacy of the mental (section 2.4). After introducing the thought experiment, I will deal with the criticism of it, the reinterpretation strategy, but I contend that it does not succeed in rebuking the thought experiment.

In order to establish anti-individualism, Burge employs the following three-step thought experiment. To

---

<sup>9</sup> I use "externalism" and "anti-individualism" interchangeably.

begin, Burge asks us to imagine a case of incomplete understanding in which an individual misconstrues (incompletely or partially understands) some notion putatively involved in the contents of some of his thoughts (step 1). In the next step (step 2), we consider a counterfactual supposition. We hold the actual individual's life history (asocially, non-rationally and non-intentionally described) and physiology constant, and suppose that the linguistic practices of the counterfactual community are such that the individual's actual incomplete understanding of the particular notion now reflects complete understanding, as determined by his (counterfactual) linguistic community (i.e., his use of the relevant term accords with the counterfactual community's linguistic conventions). The final step (step 3) involves an interpretation of the thought experiment.

In the first step of the thought experiment, Burge asks us to imagine Bert, in our actual world, whose understanding of the concept *arthritis* is partially ignorant or mistaken about the application conditions of the concept. He takes the concept to refer to inflammations of bones as well as joints. In other words his understanding of the concept is incomplete. Even if

this is so, Bert has many true beliefs about arthritis, which are correctly attributed by means of "that" clauses containing the term "arthritis." For instance, Bert believes that he has had arthritis for many years; that the arthritis in his wrists and fingers is more painful than the arthritis in his ankles, and so on. When suffering pain in his thigh, though, Bert sincerely complains to his doctor at a certain time  $t$ , "I have arthritis in my thigh." The doctor corrects him and informs him that he cannot have arthritis in his thigh, because arthritis is, by definition, a disease of the joints only. Although the belief is false, it seems that we can truly describe Bert's propositional attitude as the belief that he has arthritis in his thigh.<sup>10</sup>

In the second step Burge asks us to imagine a counterfactual situation in which Bert's physical history and intentional phenomena, individualistically described, are assumed to be the same up through the time  $t$ , but in which the term "arthritis" also applies to inflammations of the thigh. Let's call him T-Bert. The counterfactual situation differs only in that the correct, standard use

---

<sup>10</sup> The correct understanding of the issue is important to understand the debate between Burge and Davidson and I will take up the issue in detail when I am dealing with the debate.

of "arthritis" encompasses Bert's misuse. The twins have the same dispositions to assent to, or deny, the sentences "I have arthritis in my thigh."

Burge claims that in the counterfactual case we cannot correctly ascribe a belief to T-Bert with a that-clause containing our term "arthritis," because the counterfactual expression "arthritis" differs both in dictionary definition and in extension from "arthritis" as we use it. That is, "arthritis" in the counterfactual situation is not extensionally equivalent to "arthritis" in the actual situation (Burge 1979: 79). This difference, Burge claims, stems from social factors that are independent of the individual. The individual has the same physical history and intentional phenomena, individualistically described, in the actual situation as his twin does in the counterfactual situation, yet the contents of the twin's attitudes differ. T-Bert would lack beliefs involving the concept of arthritis; his belief would be said to involve the concept of, say, T-arthritis.

Given the non-indexical nature<sup>11</sup> of the twins' concepts, they have different concepts about these referents. Taken in isolation from the linguistic community, there is no way to distinguish Burt's belief and T-Burt's belief. Yet, we seem to be committed to the claim that the beliefs are different, simply in virtue of the fact that the beliefs are about different things - arthritis and T-arthritis. Given that sameness of truth-value is a necessary condition for whether beliefs are identical, the belief expressed by "I have arthritis in my thigh" in the actual situation is different from the belief expressed by a token of the same sentence type in the counterfactual community. For in the actual community the belief expressed is false, whereas in the counterfactual community it is true.

Let me explain this in detail. We ordinarily identify the contents of mental states semantically by using a complex sentence of the form "Subject A  $\Phi$ -es that  $p$ ," where " $\Phi$ " stands for a psychological verb, and " $p$ "

---

<sup>11</sup> If the concepts are indexical in nature, the twins' concepts may shift from actual situation to counterfactual situation since an indexical's referent is determined, in part, by extra-linguistic context, and therefore vary from context to context; indexicals are context-sensitive. However since the relevant concept in question is non-indexical, the difference in referents in the two circumstances entails that the twins have different concepts about these referents.

stands for a that-clause. The that-clause specifies what the mental state is about; it gives the content of the state. Thus, mental states are ordinarily understood to be content-individuated states. If beliefs and other propositional attitudes are identified and individuated by semantic content, and if semantic content is individuated in terms of their referents or truth conditions, then mental states must also be individuated in terms of their referents or truth conditions.

The specific issue of importance to us concerns the individuation of mental states, or the conditions under which mental states should count as the same or different in kind. On a very rough and practical level, mental state individuation would seem to be relatively unproblematic. Your belief that it is raining is different from my belief that I am going to play tennis, whereas your belief that 2 plus 2 is four and my belief that 2 plus 2 is four are clearly, in some intuitive sense, the same belief. The difficulties arise when we try to articulate the general conditions for beliefs being the same or different. At the most general level, there are two opposed positions with respect to the



individuation of psychological states: individualism and anti-individualism.

Now Burge's argument begins with the widely-held assumption that content clauses do not freely admit substitution of co-referring or co-extensive expressions without the possibility of changing the truth value of the containing sentence. Content clauses of propositional attitude ascriptions have traditionally been taken as a primary means of identifying a subject's intentional mental states. The motivation for this assumption is that we cannot, in general, substitute co-referring or co-extensive expressions within embedded content clauses so as to preserve the truth value of the containing sentence. Burge's line of reasoning exploits this assumption. Surely, he says, if ever co-referring expressions in oblique position<sup>12</sup> can indicate different thoughts, then it is simply undeniable that obliquely occurring expressions that are not extensionally equivalent indicate different thoughts. Burge says:

It is normal to suppose that those content clauses correctly ascribable to a person that are not in

---

<sup>12</sup> I will speak of a belief attribution's being "oblique" when the terms in a that-clause are not open to substitution by co-referential expressions *salva veritate*.

general intersubstitutable *salva veritate* - and certainly those that involve extensionally nonequivalent counterpart expressions - identify different mental states or events. (1979: 76)

This claim figures in Burge's thought experiments in that content clauses that are taken to give the attitudes of the individual, actually and counterfactually described, contain obliquely occurring expressions that are non-co-extensive in the languages in the respective communities. Burge puts this point as follows:

On any systematic theory, differences in the extension - the actual denotation, referent, or application - of counterpart expressions in that-clauses will be semantically represented, and will, in our terms, make for differences in content. (1979: 75)

On Burge's view, extensionally non-equivalent component parts of obliquely occurring content clauses clearly call for attribution of different attitudes.<sup>13</sup>

Let us now return to the thought experiment. In the final step the interpretation of the thought experiments is presented. The twins' having different mental states clearly comes from differences in their respective social

---

<sup>13</sup> According to Fodor's psychological taxonomy, mental states of the twins are the same. The psychological taxonomy should individuate the attitudes non-relationally. See Fodor 1987. We will deal with this important issue in section 2.4.

circumstances. The different *social environments* connecting the twins to different syndromes of disease necessitate that they have different beliefs with different conceptual contents. The important point to bear in mind is that even though Bert in the actual situation does not have complete linguistic mastery of a word "arthritis," he can employ the concept it expresses in his thought. Burge does not think that Bert fails to grasp the concept of arthritis. Burge writes, "[S]uch errors do not always or automatically prevent attribution of mental content provided by the very terms that are incompletely understood or misapplied" (1979: 90).<sup>14</sup>

According to Burge, "The argument can get under way in any case where it is intuitively possible to attribute a mental state or event whose content involves a notion that the subject incompletely understands ... This possibility is the key to the thought experiment" (1979: 32). In oblique position, an attitude attribution containing the term "arthritis" in the content clause can be made to Bert despite the fact that he has an incomplete understanding of the concept of arthritis. On

---

<sup>14</sup> Davidson clearly rejects this interpretation. This will be examined in section 2.3 in detail.

Burge's view, even though the individual only incompletely understands the concept of arthritis, it is still proper to say that he possesses the concept of arthritis. He is taken to have a grasp, even though it may be incomplete, of the concept of arthritis.

Of course Burge acknowledges that there are some situations in which we do not accord a subject's words their customary interpretation.<sup>15</sup> A subject, however, can be said to possess a concept just in case his use of (and dispositions to use) a term which expresses that concept are not too deviant, relative to the linguistic conventions of his community, so as to force reinterpretation of the sentences he utters (or would be disposed to utter) which contain that term. The range of "too deviant" depends on a subject's attitude; whether he is willing to have his words construed according to the socially accepted meaning, even though this requires him, in the situation in question, to accept that he said and

---

<sup>15</sup> The cases he mentions include those in which the speaker is a child, a foreigner, a speaker of a dialect, or the victim of a slip of the tongue. Here the subject either does not have full command of our standards of usage (child, foreigner), is not bound by them (dialect), or has full command but fails to manifest it because of a performance error (slip of the tongue). In each case the subject is excused from being taken at his or her word; it is assumed that the speaker did not say what they meant, or did not mean what they said except in the dialect case, where the subject did not say what we thought he said.

believed something wrong. Burge argues that the appeal to reinterpretation in the case we are discussing is not supported by the ordinary practice of mentalistic attributions. Common practice and our ordinary linguistic intuitions, he says, reveal that incomplete understanding of the meaning of a term in the common language is not incompatible with ascription of mental contents involving that term, literally interpreted, which is to say interpreted in accordance with common linguistic practice.

Burge considers two general strategies for reinterpreting the thought experiments, and criticizes these methods of reinterpretation. The first strategy he considers for reinterpreting the thought experiments is the attempt to motivate a non-literal reading of the sentence that the individual uses to express his belief, which directly displays the subject's incomplete understanding. The second general strategy<sup>16</sup> for

---

<sup>16</sup> In his 1979 Burge deals with four methods that are supposed to provide an alternative interpretation of the thought experiments. The first method for reinterpreting the thought experiment that Burge considers involves an appeal to *de re* beliefs. On Burge's view, a *de re* belief is a belief which relates an individual to an actual object. The second method of reinterpreting the thought experiment holds that in cases of incomplete understanding, the content of the individual's attitude is indefinite. The third method is called "object-level" method of reinterpretation, of which Burge says, "One is to attribute a notion that just captures the misconception, thus replacing contents that are apparently false on account of the misconception, by true contents" (1979: 93). The last, closely

reinterpreting the thought experiments that Burge considers attempts to sever the connection between the contents of the subject's attitudes and the proposition expressed by the sentences which are used to attribute the contents.

The problem these reinterpretation strategies present for Burge's argument is that if the sentence the subject uses contains words that we know he doesn't fully understand, then that sentence should not be understood literally. If this were the case, it would not be correct to say that the subject's belief is false. And recall, Burge's grounds for distinguishing the actual individual's belief from the counterfactual individual's belief is that they differ in truth value.

Burge's criticism of these methods of reinterpretation is based on two general claims. First, Burge says, the methods fail to account for the practice of ordinary mentalistic attributions (what we typically say and do when we catch others using words incorrectly). And second, the reinterpretations urged by the methods

---

related, method of reinterpreting the thought experiments, the "metalinguistic" method, proceeds from the claim that the individual's incomplete understanding is more accurately described as a metalinguistic error. This method attempts to account for the individual's misuse of the particular term.

are not supported by what the individual would say and do when he realizes that he had been using the particular term incorrectly.

To begin with the first of these claims, Burge's view is that we do not typically (in ordinary practice) search for true object-level contents, nor do we ordinarily suppose that all of the individual's attitudes involving the misconstrued term involve reference to expression at the metalinguistic level. Burge's second general claim is that the metalinguistic and the object-level methods of reinterpretation are committed to a highly implausible account of how the individual would react when he discovers that he had been using a term incorrectly. When, for example, the subject learns what arthritis is, he does not, Burge contends, typically respond by saying that his views have been misunderstood. Rather, the individual is typically willing to revise his use of the term on the authority of an expert or a reliable source. Moreover, the individual typically admits that the belief he had expressed by saying "I have arthritis in my thigh" was false. This suggests that the individual intended to have his words taken literally.

## 2.2 DAVIDSON'S EXTERNALISM

In this section I will introduce Davidson's so-called triangular externalism in order to clearly see the debate between Burge and Davidson in the next section.

Davidson's own brand of externalism differs in relevant ways from what has been generally called "externalism," particularly that of Tyler Burge. Davidson does not rely on Twin-Earth thought experiments to establish his variety of externalism. Rather, Davidson motivates his triangular externalism by appealing directly to facts about language learning and considerations about how we interpret words and languages with which we are unfamiliar. Davidson thus thinks that the thesis of the external individuation and constitution of thoughts is a direct consequence of the way the basic connection between words and things or thoughts and world is established.

Davidson agrees with Burge that externalism is not restricted to natural kind terms, but extends to language and thought generally. Davidson also accepts the externalist thesis that our mental contents are externally determined. He concurs with Burge that two



thinkers may be alike in all relevant physical respects and yet differ in their ordinary psychological states, for instance, they may mean quite different things with the word "water" (Davidson 1988).<sup>17</sup>

Davidson, however, doesn't accept the particular way in which Burge thinks external factors are relevant to the individuation of content. Davidson provides three main reasons why he rejects Burge's social externalism (1991: 198-9): first, it seems to be unintuitive to elicit speaker's meaning from an elite usage; second, if speaker's meaning is determined in terms of what other people in the community would mean by the same words, then first person authority necessarily lapses; third, Davidson distrusts thought experiments because they are impractical.

Davidson thinks it is wrong to hold the idea that as speakers we have an obligation to the language, or the community, or our audience, to speak according to some standard. Whether or not Burge actually holds this idea, this is the way Davidson interprets Burge. Within the

---

<sup>17</sup> Davidson has not explicitly argued why he would not allow local supervenience. He seems to reject local supervenience after taking Burge's thought experiments seriously; the explicit expression that he does not allow local supervenience first appeared in his 1987. This important issue will be emphasized when I deal with the Sunburn Argument.

Davidsonian picture, such obligations, though they sometimes exist, are irrelevant to communication, because the crucial point for Davidson is for the speaker necessarily to intend to speak in a way that will be understood along the intention. For Davidson the only interesting concept of meaning must derive from cases of successful communication. Successful communication, Davidson claims, cannot be defined in terms of shared meanings, practices or conventions.

The problem with Burge's social externalism, according to Davidson, is that it allows public conventions to determine content. This seems to make content independent of the speaker's intentions. Davidson's claim is that intentional states, such as belief and desires, are individuated by causal relations to objects in the world. In determining the concepts and thoughts of an individual, Davidson rejects Burge's externalism and the normative role of the linguistic community. His reason for this is that what determines the possession of a concept is not membership in a particular linguistic community, but the acquisition of a disposition through causal contact with objects and events in a social setting. On Davidson's view the

differences in meanings and psychological states, discussed by Burge, result from the history of causal relations between the individual thinker, others with whom he communicates, and the natural environment (Davidson 1991: 203-204).

The triangle between teacher, learner, and environment is basic to learning a language and to interpreting the thoughts and meanings of others.<sup>18</sup> While Davidson agrees that two thinkers may be in type-identical physical states and still think different "water" thoughts, he emphasizes that there is a difference in the causal history of the respective thoughts, e.g., the two thinkers learned the word form "water" in different natural and social settings.

[The basic connection between words and things] is established by causal interactions between people and parts and aspects of the world. The disposition to react differentially to objects and events thus set up are central to the correct interpretation of a person's thoughts and speech. If this were not the case, we would have no way of discovering what others think, or what they mean by their words. The principle is as obvious and simple as this; a sentence someone is inspired (caused) to hold true

---

<sup>18</sup> The social and non-social aspects of Davidson's externalism are not independent of one another in that both result from the way the basic connection between words and things and thoughts and speech is established in the triangulation of speaker, others with whom she interacts, and objects and events in the environment.

by and only by sightings of the moon is apt to mean something like 'There's the moon'; the thought expressed is apt to be that the moon is there; the thought inspired by and only by sightings of the moon is apt to be the thought that the moon is there ... Not that all words and sentences are this directly conditioned to what they are about; we can perfectly well learn to use the word 'moon' without ever seeing it. The claim is that all thought and language must have a foundation in such direct historical connections, and these connections constrain the interpretation of thoughts and speech. Perhaps I should stress that the argument for this claim does not rest on intuitions concerning what we would say if certain counterfactuals were true. No science fiction or thought experiments are required. (Davidson 1987: 29)

Davidson thus traces the individuation of meanings, concepts and mental states like beliefs to patterns of causal interactions in the triangulation of the individual, other speakers with whom he or she interacts, and objects and events in the world. These patterns of causal interactions are not determined by the world itself or by the norms of a linguistic community, but by the contextual and social use of words to apply to objects and events.

Davidson's triangular externalism differs from Burge's anti-individualism with respect to how the contents of propositional attitudes are externally individuated. While Davidson agrees with Burge that

social factors play a role in the external individuation of mental contents, he locates the social factors involved in "the causal nexus that includes the interplay between persons and the rest of nature" (Davison 1991: 201).

### **2.3 EXTERNALISM AND TOKEN IDENTITY: BURGE AND DAVIDSON**

Some philosophers give an argument, claiming that if our mental states do not supervene on properties intrinsic to our bodies, then all versions of psychophysical identity theory seem to be threatened. This was first pointed out by Burge (1979), among others. Davidson does not think that his AM is open to the threat from externalism. In this section I will examine the debate between Davidson and Burge on the issue whether Davidson's AM is compatible with content externalism. The purpose for looking at the debate is, first, to strengthen the claim that AM is wrong, and therefore that NCC should be rejected, and second, to help elucidate content externalism in general which has a lot of implications on the issues in the following chapters.

I will present Burge's argument against the token identity thesis and Davidson's response, the so-called Sunburn Argument, to Burge. Burge's strategy is to show that AM is incompatible with content externalism. I argue that Davidson does not succeed in showing that AM is compatible with content externalism.

The following is Burge's argument against the token identity thesis. Take any physical event-token *p* correlated with a subject while she thinks that arthritis is a painful disease: *p* is a *plausible candidate* for identification with a mental event *m*, thinking that arthritis is a painful disease, and is specifiable by physical sciences such as physics, chemistry, and neurophysiology. Burge's thought experiment shows that it is possible for a subject to think a thought with different contents, *m\**, even though the same event-token *p* occur in the subject's body: for example, in the counterfactual situation the same event-token *p* occurs without her having any thought, *m*, that arthritis is a painful disease; *p* could occur with her having the thought, *m\**, that T-arthritis is a painful disease.<sup>19</sup>

---

<sup>19</sup> Burge says that this possibility is not entailed by his thought experiments, even though it is strongly suggested (Burge 1993: 105).

However any occurrence of thought could not have a different content and be the very same token event: a thought with the intentional content  $m$  and a thought with the intentional content  $m^*$  cannot be the very same event-token. Therefore it is not the case that  $p$  is  $m$  because  $p$  could occur without  $m$  occurring: the same event-token  $p$  is not the subject's thought that arthritis is a painful disease (1993: 104-113; 1979: 110-111).<sup>20</sup>

From the argument against the token identity thesis, Burge rejects NCC. Burge claims that we do not know and cannot know *a priori* that causal statements entail the existence of strict laws. Unless mental causation is just physical causation there is no reason to think that it would interfere with physical processes. To think this is already to think of mental causation on a physical model,

---

The reason for Burge to say that it's not entailed can be seen in his 1989 paper:

[The anti-individualistic] conception does not entail that two individuals' mental kinds might differ while relevantly corresponding brain states and events remain type-identical. Failure of supervenience of an individual's mental kinds on his neural kinds follows only if relevant differences in the environment do not necessitate differences in the individual's underlying brain states. (1989: 305)

Now we can conceive that Burge's arthritis thought experiment is the case that the relevant social difference does not necessitate differences in the individual's underlying brain states.

<sup>20</sup> Here one of the premises was that  $p$  is a *plausible candidate* for identification with a mental event  $m$ , but we found that the premise is false:  $p$  cannot be  $m$ .

which there is no reason to do. Interference would be surprising. So non-interference is in no need of explanation in ontological terms.

In the face of Burge's attack on this matter, Davidson presents the Sunburn Argument. The argument tries to show that there is no incompatibility between externalism and AM. The argument goes as follows:

[The Sunburn Argument]

- I Two individuals' mental kinds might differ while relevantly corresponding brain states and events remain type-identical.
- II Identifying a condition as sunburn does not mean that a sunburn is not a state of the skin.
- III Mental states are like sunburn in the above respect.

Therefore,

- IV Mental states can be token-identical with physical states with a person.

Just as identifying a condition as a sunburn does not mean that a sunburn is not a state of the skin, so identifying mental states by external factors does not entail that they are not states of the head. Davidson claims that though the sunburned skin and the skin burned by a sunlamp may be indistinguishable, still it does not follow that two states (sunburn and sunlamp-burn) are the same. It is because one state is from the sun and the



other sunlamp. With regard to this, mental states are just like sunburn. He writes:

There may be no physical difference between being sunburned and being burned by a sunlamp, but there is a difference, since one state was and the other was not caused by the sun. Psychological states are in this respect like sunburn. (1988: 49)

Even if we need to appeal to the extrinsic causes of the respective skin conditions in order to individuate them as being sunburn and sunlamp-burn, this doesn't mean that they aren't conditions of the skin. To say that a condition of one's skin - say, a sunburn - supervenes on what caused it, does not entail that the condition is not "in" one's skin. This point is the gist of the Sunburn Argument. Davidson claims that the alleged difficulty stems from unquestioned assumptions, namely, "If a thought is identified by a relation to something outside the head, it isn't wholly in the head. (It ain't in the head.)" (1987: 31) Mental states can be regarded to be physical states of a person, yet to be causally dependent on factors external to that person's body. The externalist, Davidson says, can thus claim that mental states are identical with physical states of a person, but that they are causally dependent on factors outside

the person's body. Davidson writes, "This is enough to show that an appreciation of the external factors that enter into our common ways of identifying mental states does not discredit an identity theory of the mental and the physical" (1987: 31-2).

Now I attack Davidson's Sunburn Argument for the following reasons, which are closely interrelated with each other. First, it is not clear why Davidson claims the first premise, the failure of local supervenience in the Sunburn Argument. Second, the analogy does not work, therefore, the third premise is wrong. Third, and the most important, if the Sunburn Argument works we lose global supervenience. Before we turn to the discussion of the three reasons, let me emphasize on three points. First, Davidson does not use counterfactual situations to establish his externalism. Second, Davidson rejects Burge's first step of thought experiment. And third, Davidson rejects Burge's Twin-Earth thought experiments in general. I will mention the last two points in detail.

Let me explain the last point first. It is clear that Davidson does not (and of course, need not) follow, in a step-by-step way, Burge's argument against the token identity thesis in order to show that AM is compatible

with content externalism. For example, Davidson says, "I have a general distrust of thought experiments that pretend to reveal what we would say under conditions that in fact never arise" (1991: 199). He does not buy the specific procedure that Burge takes, though he favors some kind of externalism as we saw in the previous section. The following passage shows that Davidson rejects Burge's Twin-Earth thought experiments in general:

[I]f Burge is right, then whenever a person is wrong, confused, or partially misinformed, about the public meaning of a word, he is wrong, confused, or partially misinformed about any of his beliefs that are (or would be?) expressed by using that word. Since such 'partial understanding' is 'common or even normal in the case of a large number of expressions in our vocabularies' according to Burge, it must be equally common or normal for us to be wrong about what we believe ... I must reject some premise of Burge's<sup>21</sup>. I agree that what I mean and think is not 'fixed' (exclusively) by what goes on in me, so what I must reject is Burge's account of

---

<sup>21</sup> One of the reasons Davidson thinks he should reject Burge's social externalism is that it is not compatible with the presumption that we have first person authority. So for example, Davidson claims in another place that "there is a conflict between Burge's social externalism, which ties a speaker's meaning to an elite usage he may not be aware of, and first person authority." (1991: 199). However, this reason is not persuasive. Burge, in his 1988 paper, actually argues, I believe successfully, that there is no conflict between anti-individualism and first person authority.

how social and other external factors control the contents of a person's mind. (Davidson 1987: 26-27)

Now let's take a look at the second point that Davidson rejects Burge's first step of thought experiment. In the "arthritis" thought experiment, Burge claims that Bert's incomplete linguistic mastery of a word "arthritis" does not prevent him from employing the concept it expresses in his thought. As we already saw, Burge does not think that Bert fails to grasp the concept of arthritis. Burge believes that the doctor and patient can share beliefs like the belief that arthritis is a painful disease, and thus can share the concept of arthritis. They can do this even though the patient is mistaken about some fundamental features of arthritis and has vastly less background knowledge than the doctor. However, Davidson rejects this construal of Burge's explanation about incomplete understanding. According to Davidson, there is a relevant difference in the thoughts between Bert and a doctor who has a full mastery of the concept *arthritis* (1987: 27). Davidson does not say very much on this except appealing to holism about belief and the uncontroversial point that Bert would associate arthritis with different background beliefs and

inferences from someone who knows that arthritis can occur only in joints. He holds that the error is a metalinguistic one about the dictionary meaning of the word "arthritis." The point, however, is that there is a difference between the "concept" or its linguistic counterpart "translational meaning" and the "the conceptual explication" or "explicational meaning" (Burge 1989: 180-7). The latter is subject to correction or confirmation by empirical consideration of the referents. Burge thinks that Davidson makes a mistake in failing to recognize the difference between being able to understand well enough (the former), and being able to give a correct explication (the latter).

Now we are in a position to attack the Sunburn Argument. In the previous section I mentioned that Davidson's own brand of externalism differs from that of Burge. Davidson affirms the idea that mental states supervene globally on physical states of a person and factors in the environment.

[S]ubjective states are not supervenient on the state of the brain or nervous system: two people may be in the same physical state and yet be in different psychological states. This does not mean, of course, that mental states are not supervenient on physical states, for there must be a difference

*somewhere* if psychological states are different. The interesting physical difference may not be in the person; like the difference between water and twater, it may be (we are supposing) elsewhere. (Davidson 1989:61-62)

Davidson's point of using the "Sunburn" analogy is that the conditions such as sunburn similarly supervene on physical properties of the skin and the extrinsic causal conditions. However, he does not give an argument why he accepts the result of Burge's thought experiment, namely the first premise of the Sunburn Argument. Davidson has not explained why he rejects local supervenience<sup>22</sup>.

In several places he says that two people may be in the same physical state but differ in what they think. This is the first premise of the Sunburn Argument. Davidson just accepts the result of Burge's thought experiments. But what is Davidson's argument for the failure of local supervenience since he generally distrusts Burge's thought experiments? Without a counterfactual supposition, or a science fiction if I use Davidson's terminology, we cannot imagine, practically speaking, again if I use Davidson's terminology, the

---

<sup>22</sup> Davidson's triangular externalism is not enough to establish the failure of local supervenience.

situation that two people are the same in all physical respects.

The above discussion naturally leads to the second reason against the Sunburn Argument. The Sunburn Argument concentrates on the analogy of sunburned states with mental states. However, we don't see any convincing reason to take the analogy. On the other hand we have every reason to reject the analogy. Burge thinks, I believe wrongly, that Davidson's Sunburn Argument shows that the difference in causal histories between the twins would necessitate a difference in the physical states of the twins. Therefore Burge thinks that the Twin-Earth cases would never illustrate a case in which the internal physical states of the twins would be the same while the mental states differed. The following remark by Burge confirms my interpretation. After he asks whether it makes sense to individuate brain states depending on causal histories, he says:

There certainly are physical differences between actual and counterfactual situations in the relevant thought experiments. The question is whether there are always physically different entities that are plausible candidates for being identical with the different mental events or state-instances. The different physical causal histories are not plausible candidates. These histories do not have

the same causes or effects that the relevant mental events (states) do. Moreover, it is doubtful that relevantly described causal histories instantiate explanatory natural kinds in any of the physical sciences. ... What is objectionable about this view is that it makes the individuation of brain events depend on matters that are irrelevant to the physiology of the brain. (1993: 106-107)

He wrongly believes Davidson claims that the different causal histories of sunburn and sunlamp-burn would make a difference in the physical entities. Even though Burge is wrong on this, his argument still works: The Sunburn Argument does not save Davidson's AM.

Now let us return to the second reason again. If the case of sunburn and sunlamp-burn is a case that shows that local supervenience fails as Davidson thinks it does, then there is no physical difference in the persons that have sunburn and sunlamp-burn. Then the physical difference should be elsewhere. Davidson says, "The interesting physical difference may not be in the person; like the difference between water and twater, it may be (we are supposing) elsewhere" (Davidson 1989:62). However now it is difficult to imagine where is the interesting physical difference if not in the person in the case of sunburn and sunlamp-burn. If we accept Davidson's Sunburn Argument, the difference in causal histories between the



twins would necessitate a difference in the twins' mental states while physical states of the twins plus the physical world are the same. In order for the analogy to work the internal physical states between sunburned skin and the skin burned by a sunlamp would be the same while the interesting physical difference would be elsewhere, somewhere in the physical world: even though local supervenience fails, global supervenience should work.

This discussion now leads to the third reason to argue against the Sunburn Argument. How is the physical world somewhere else different? The difference in causal histories between sunburn and sunlamp-burn would never necessitate a difference in the physical world. A disastrous result! Even global supervenience fails in the Sunburn Argument. Let's take a look the following remark of Davidson:

People who are in all relevant physical respects similar can differ in what they mean or think, just as they can differ in being grandfathers or being sunburned. But of course there is something different about them, even in the physical world; their causal histories are different, and they are discrete physical objects. We are therefore free to hold that people can be in all relevant physical respects identical (identical in 'necktie sense') while differing psychologically. (1989)

Now this remark is very confusing. As I said without supposing the counterfactual situation there is no point of introducing two physically similar people. Davidson says that he is able to argue for the failure of local supervenience, i.e. the possibility that two thinkers may be in type-identical physical states and still think differently, without requiring the counterfactual thought experiment. However, the whole point is that we cannot just assume in *this* world that there are two people exactly in the same physical states. More importantly, in this world there is just *one* global supervenience base. Of course sunburn and sunlamp-burn have different causal histories, but sunburn and sunlamp-burn has the same, one, global supervenience base. Then global supervenience fails in the Sunburn Argument.

#### **2.4 CAUSAL EFFICACY OF EXTERNALLY INDIVIDUATED MENTAL CONTENT: BURGE VS. FODOR**

In this section I begin examining specific arguments against the causal efficacy of externalistic mental states. This is a question about how propositional attitude states, externally individuated, can enter into

true causal explanations of action. I will support the claim that beliefs and other mental states with widely individuated intentional contents play genuine causal roles in virtue of their contents in psychological causal explanations of behavior. I present Fodor's challenge, by examining his *Cross-Context Argument*, an argument that externalism eliminates the causal relevance of the mental, and I provide some possible responses for nonreductive physicalism.

The Cross-Context Argument is designed to show that externally individuated contents are not causally efficacious. In his 1987, Fodor argues that we would judge that the effects of distinct wide contents in the same context would be the same. He says, "[I]dentity of causal powers has to be assessed *across* contexts, not *within* contexts" (1987: 35). To individuate across contexts is to make judgments of sameness and difference while keeping contexts constant. The following is the good example to illustrate this point. It is true that as the effect of my utterance "water" I get water and the effect of my Twin's saying "water" my Twin gets T-water. But, Fodor claims, these effects of our causal powers only differ because they occur in different contexts, and

we cannot conclude anything about the sameness or difference of our causal powers based on differences in effects that occur in different contexts. The criteria for determining the identity of causal powers are as follows:

(a) if his utterance (/thought) had occurred in my context, it *would have had* the effects that my utterance (/thought) did have; and (b) if my utterance (/thought) had occurred in his context, it *would have had* the effects that his utterance (/thought) did have. For our utterances (/thoughts) to have the same causal powers, both of these counterfactuals have to be true. But both of these counterfactuals *are* true, since (for example) if I had said "Bring water!" on Twin-Earth, it's XYZ that my interlocutors would have brought; and if he had said "Bring water!" here, his interlocutors would have brought him H<sub>2</sub>O. (Fodor 1987: 35)

The above pair of counterfactuals is the tool for assessing across contexts. Wide content differences, Fodor argues, would not make a difference to causal powers, which means wide content would not count as causal powers in science. This is because what Fodor considers a general principle in science is that no property counts taxonomically unless it affects or makes a difference to causal powers.

Fodor concludes that externalistic contents do not pass this cross-context test. If we judge that our causal

powers would have the same effects in all the same contexts, then our causal powers are the same. If the cross-context test shows that causal powers are the same even when wide contents differ, then wide contents differences are causally irrelevant.

Burge responds to the argument by saying that The Cross-Context argument does not show that widely individuated properties do not have causal powers. Burge argues that the value of the test depends entirely on which contexts are considered relevant. We can only infer to sameness of causes from sameness of effects in contexts where a difference of causes could make a difference if there is one. Burge says:

There could be a device that traced the histories of individuals, recording whether they had been in causal contact with [water]. Such a device could bring [water] to an individual with such a causal history when he made the sounds "Bring [water]" - and not otherwise. In such a context, A would have different effects from [Twin-A]. ... [T]here is a possible context in which the twins' acts produce different effects. Unless some restriction is placed on admissible contexts, Fodor's test will count any two individuals with any differences at all in their physical histories as having different causal powers. (Burge 1989a: 311)

As Burge points out, ruling out contexts where causal-historical properties might make a difference in effects is question-begging.

Burge's second point is that Fodor's Cross-Context test is insensitive to the environmental background against which the individuals' psychological states are type-individuated. The individuals' causal powers are relative to each science and its explanatory concerns. Let me explain this by using Burge's cases of pumping blood and pumping waste. If a heart were to replace a physically homologous organ whose function is to pump waste, the heart would have the same physical effects as its physically homologous counterpart. But the heart and the homologous waste-pump would not have the same causal powers as typed by physiology. Burge says that it is ludicrous from this fact to argue that:

[T]he heart and its counterpart have the same causal powers *as typed by physiology* and that there is no difference in kind. From the point of view of some sciences, the two entities would indeed count as type identical. But the physiological differences are patent. Physiology recognizes causal powers of the heart which are exercised in its functionally normal environment. ... But these environments are irrelevant to the scheme of kind individuation that physiology actually uses. Fodor's test is insensitive to this dependence of many special sciences on a normal environment for picking out

those causal powers that are relevant to an explanatory typology in those sciences. (Burge 1989a: 312-313)

Similarly, the conception of causal power in psychology is taken "not from some model drawn from the other sciences, but from the explanations that psychology provides" (1989a: 316). The example of the heart and the organ that pumps waste provides the case where they have same causal powers as typed by physics but they have different causal powers seen from psychology. What this means is that the twins with the same causal powers as typed by, for example, neurophysiology have different causal powers seen from a higher-level special science.

In this section we saw that if we are to find a genuine explanatory role for content, we must accept the fact that widely individuated, relational properties can have causal relevance. We saw Burge's solution as to how we may see widely individuated propositional attitude properties as playing crucial explanatory roles in genuinely causal explanations.

## **2.5 CONCLUSION: REJECTING NCC**

In this chapter I have dealt with a tension that arises from content externalism. We also examined the debate between Burge and Davidson to strengthen the claim in Chapter One that AM is committed to the epiphenomenalism of the mental, and therefore that NCC should be rejected. By examining the debate between Burge and Fodor we saw that there is no *a priori* reason why the so-called "wide" contents do not or cannot play causal roles in psychological explanations of behavior, and showed how they might do so by noting that wide contents are among the properties we ordinarily cite to explain our behavior.

Fodor basically argues that individuals cannot have different causal powers without their having different brain states. The motivation for holding this, according to Burge, is that he believes "physiological processes are where the "real" causation in psychology goes on" (Burge 1989a: 306). Burge calls this a crude version. The most deeply imaginative version of this is executed, according to Burge, by Davidson's Nomological Character of Causality. Burge says that:

Davidson holds that attribution of causal relations entails commitment to a certain sort of explanatory



law, a sort of law that has properties ... that one cannot reasonably expect the principles of psychology to exhibit. Mind-body causation is then interpreted in the light of this assumption. Such causation is held to fall under purely physical laws (Burge 1989a: 317-318).

Burge claims that there is no *a priori* reason to think that way, therefore it is an empirical question. Burge claims that "One cannot know *a priori* that every causal relation, regardless of domain, must fall under laws that have any particular form," and "what counts as a law is filled out partly through scientific practice" (1989a: 318).

Widely individuated properties can have causal relevance in that the explananda of psychology are taken to be behavioral events under relational descriptions. An issue exists as to whether scientific psychology ought to take behavioral events under intentional descriptions as its explananda; but it seems perfectly clear that commonsense psychology is precisely in the business of explaining individual bits of behavior intentionally described. The question was whether explanations of intentionally (relationally) described behavior, explanations making use of relational propositional attitude properties, are genuinely causal explanations,

given the fact that wide content can vary without affecting causal powers. Burge shows how it works.

The manner in which propositional attitude properties manage to play non-superfluous causal/explanatory roles is just the following: adverting to such properties enables us to give causal explanations of facts (intentionally characterized facts) that we could not otherwise explain. The internal conceptions that causally explain our actions may be intrinsic to our brains in that such internal conceptions do supervene upon internal microstructure. But, in order to characterize those internal conceptions for purposes of explanation of action, namely in order to speak of mental content at all, we must ascribe relational properties to one another. According to Burge, mentalistic explanation is a key to understanding mental-physical causation. Burge claims:

Understanding psychological causation is at least as dependent on what sorts of explanations we achieve in psychology, and how they are related to explanations in the biological sciences, as it is on any antecedent conception of causation. It is therefore an open question whether it will ever be illuminating and correct to count relations between neural events (tokens) as revealing the nature of causal relations involving intentional psychological events. (1989a: 318)

Causal explanations of action must, therefore, make reference to relational properties, properties which do not affect the causal powers of internal states, but which are nevertheless explanatorily indispensable.

### CHAPTER 3

#### THE EXCLUSION ARGUMENT AND THE PRINCIPLE OF THE NOMOLOGICAL CHARACTER OF CAUSALITY

The Exclusion Argument is designed to show that nonreductive conceptions of the mental face the serious problem of producing an account of mental causation which does not render the mental epiphenomenal.<sup>23</sup> Recall that there is more than one problem of mental causation. One problem is the problem presented by Davidson's AM. Another problem is one presented by the failure of mental content to supervene on the physical. A third is the problem presented by the Exclusion Argument. The exclusion problem is arguably the only one which applies to any kind of mental property or state. The problem of externalism just applies to representational or contentful states, since it is only regarding these states that local supervenience is supposed to fail. The same can be said about the problem of anomalism. As Davidson himself states, the thesis of mental anomalism

---

<sup>23</sup> Kim sometimes focused exclusively on Davidson's AM. Kim, however, thinks that all nonreductive accounts of the mental face serious problems when it comes to telling a coherent story about mental causation. See Kim (1998).

covers just those states governed by considerations of rationality, namely propositional attitudes.

In this chapter I will argue against the Exclusion Argument. The unsoundness of the Exclusion Argument, however, does not save Davidson's Anomalous Monism. In section 3.1 I will show that NCC is a doctrine which can in fact be questioned. I mentioned in Chapter One that the assumption of NCC in AM is responsible for the problem of mental causation. I claim the root for the unsoundness in both the Exclusion Argument and AM results from the same incorrect intuitions: NCC. Since there is no *a priori* reason to accept NCC and there is plenty of evidence showing that NCC is actually a dubious principle, I argue against NCC.

Before I advance the claim that dealing with the Exclusion Argument casts sufficient doubt on NCC to license its rejection, I will formulate, in section 3.2, what I consider to be the most plausible version of the Exclusion Argument, Kim's argument, which seems to be an insurmountable problem for the causal efficacy of the mental for nonreductive physicalism and will outline the precise structure of the argument. In section 3.3 I will argue against the Exclusion Argument by showing that the

causal relations between mental properties do not depend on causal relations between microproperties that realize them. There is little reason to understand causation in the manner required to make the argument work.<sup>24</sup>

One of the main principles that the Exclusion Argument is using is the Causal Inheritance Principle (CIP). I will show in section 3.4 that the rejection of the argument is followed by the rejection of CIP, which says that a mental property, realized in virtue of a physical realization base, has no new causal powers beyond the causal powers of the physical base. This is important because the rejection of CIP entails the rejection of NCC.<sup>25</sup> If we take content externalism seriously, and of course we should, CIP is literally false. Contrary to the claim of CIP, mental properties do not inherit their causal powers from the properties that realize them. I will conclude this chapter by briefly considering our explanatory practice.

---

<sup>24</sup> In Chapter Four I will respond to Kim's challenge by using Baker and Burge's proposal to think about the causal efficacy of specific properties in the context of established scientific and commonsensical explanatory practices.

<sup>25</sup> The close relation between CIP and NCC will be pursued fully in Chapter Six.

### 3.1 SOME INTUITIONS AGAINST NCC

The Exclusion Principle says that there is no more than one complete and independent cause of any event. However, I point out that it seems to be unjust to single out one level of description as the "real" explanatory level, leaving others out as pseudo-explanation. We do not need to view the options as an exclusive choice. For it is possible to have different descriptions of the same phenomena. Indeed, this is what Davidson has famously argued for in his AM. Mentalistic descriptions can refer to the very same phenomena picked out by physical descriptions. Mental explanations and neurophysiological explanations are not in competition, but are rather alternative modes of picking out the very same patterns of the world around us. What shows that these explanations are not in competition is the claim that mental events just are physical events and that causation is extensional in nature; that is, that how we describe things has no impact on their causal efficacy. The difference between mental and physical explanations has to do with how mental states are picked out. It is worth noting that, at the very least, it does answer the

question of how intentional and physical explanations relate to one another: they are different ways of looking at the same phenomena, for mental events are brain events on AM.

However, someone might object to this by arguing that we are left with a gap in our explanatory practices. We still need an answer to the following question: "What do neurophysiological explanations have to do with psychological explanations?" In other words, we are left with a mystery if we leave a sharp gap between intentional explanations and physical explanations. Why does anything that happens to the brain have any effect on the mind, and vice versa? Given the fact that mental explanations and neurophysiological explanations have proven themselves successful at picking out causal relations, how do such explanations relate to one another?

At this point I start to take a position against Davidson. The worry is that we don't have any clear explanation of the gap between intentional explanations and physical explanations. However, the requirement that we have a clear explanation of the gap seems to follow from the Principle of the Nomological Character of



Causality. NCC states that events related as cause and effect fall under strict deterministic laws; if a singular causal statement connecting two events *a* and *b* is true, then there must be a causal law connecting them, namely, there must be physical descriptions of those mental and physical events such that the singular causal statement connecting those events under those descriptions instantiates a causal law. However, if we reject NCC, we don't need to worry about finding some explanations relating intentional explanations to physical explanations, because commonsense psychology is precisely in the business of explaining individual bits of behavior intentionally described.

There are reasons to doubt NCC. Our mental states can play a causal role without thereby being reducible to the language of a scientific theory. The central point is that the singular causal statements we invoke in action explanations are not in need of any appeal to regularity or law, but are themselves legitimate. The motivation for the demand for laws in action explanations stems at least in part from the fact that the laws cited in explanations are the laws that subsume events in naturalistic causal relations. By rejecting the idea that causal explanation

is causal because it is grounded in natural causal relations, the motivation for requiring laws in explanations disappears.

It is by recognizing the legitimacy and importance of the sorts of singular causal statements that are involved in the attribution of mental states to ourselves and others that the epiphenomenalist worries about the mental can be ruled out. In addition to this fact, many accepted psychological causal explanations, like many explanations in general, do not cite laws. If the considerations outlined above are correct, then they seem to provide intuitive reasons to doubt NCC.

### **3.2 THE EXCLUSION ARGUMENT**

In this section I will formulate what I consider to be the most plausible version of the Exclusion Argument, Kim's argument, which seems to be an insurmountable problem for the causal efficacy of the mental, given nonreductive physicalism, and I outline the precise structure of the argument.

What has been perhaps the most influential treatment of the exclusion problem, namely that in Jaegwon Kim's

papers "Explanatory Realism, Causal Realism and Explanatory Exclusion" and "Mechanism, Purpose and Explanatory Exclusion", is presented, as it is indicated by the titles, primarily in an explanatory way. Kim considers that both the explanatory and the causal considerations are roughly equivalent, probably the epistemological and ontological sides of the same coin. This is why he uses several times the expression "causal/explanatory exclusion" and also why, when he is using the explanatory principle, he refers in general to causal explanation. Thus, in contexts in which his main worries are related to causation he uses the causal formulation, and in contexts in which he deals with explanatory issues he prefers the explanatory one. Kim says: "It seems to me that the case for explanatory exclusion is most persuasively made for causal explanations of individual events" (1989a: 250), and proceeds to make his case accordingly. When it is argued that causal explanations exclude each other, reasons are given in terms of "sufficient causes", "causal links" and "causal overdetermination." This is particularly important since exclusion is defended by showing the

implausibility of alternative possibilities, and such possibilities are all causally formulated.

Kim's worry is whether the causal/explanatory role of mental properties can be regarded as truly autonomous and is not free-riding on the underlying physical mechanism. And he says it can't be. Kim's challenge to mental causation within the framework of nonreductive physicalism, Kim's Exclusion Argument, can be reconstructed in the following manner. Let us assume the following:  $M_1$  causes  $M_2$ ,  $M_1$  and  $M_2$  are realized by physical states,  $P_1$  and  $P_2$ , respectively; and  $M_1$  is not identical to  $P_1$  and  $M_2$  is not identical to  $P_2$ . I am using the terms 'the mental' and 'the physical' to refer to particular instances of the mental and physical properties, respectively.<sup>26</sup> Now the following is the reconstruction of Kim's causal/explanatory Exclusion Argument:

[Causal/Explanatory Exclusion Argument]

I        There is downward causation<sup>27</sup> by irreducible mental properties.

---

<sup>26</sup> I will talk of mental properties, like desiring that  $p$ , and their instantiations, James' desiring that  $p$  at time  $t$ . I will speak about the instantiation of mental properties by persons. When I speak of properties, I will usually mean property instantiations, as the context will make clear.

<sup>27</sup> The case of mental-to-physical causation is an example of downward causation. According to his Supervenience Argument (Kim 1998), for

II If there is downward causation by irreducible mental properties, there are two distinct nomologically sufficient conditions of a single event.

III For a single event, there are not two distinct nomologically sufficient conditions.

Therefore,

IV The irreducible mental properties are not causally efficacious.

This is Kim's famous Exclusion Argument.

Suppose  $M_1$ , which is not reducible to any physical properties, causes  $M_2$ . Kim invites us to ask "Why is this instance of  $M_2$  present?" (Kim 1993b: 351) Kim says two answers can be given to the question: on the one hand, the instance of  $M_2$  is there because of the instance of  $M_1$ 's causing the instance of  $M_2$ ; on the other hand,  $M_2$  is there because the instantiation of  $P_2$  realized  $M_2$ .

According to Kim, we need to explain this situation because it creates a tension. Kim says the only coherent answer to this tension is to suggest a kind of "downward causation" from the mental to the physical, from  $M_1$  to  $P_2$ . In other words,  $M_1$  caused  $M_2$  *by causing  $P_2$ ,  $M_2$ 's physical realization base*. From this consideration Kim elicits the following principle:

---

example, mental-to-mental causation is possible only if mental-to-physical causation is possible.

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

The gist of this principle is that whenever there is mental to mental causation, there is downward causation: "What these reflections show is that within the stratified world of nonreductive physicalism ..., "same-level" causation can occur only if "cross-level" causation can occur" (Kim 1993b: 353). Kim says that most nonreductive physicalists should accept this principle.

The next principles we need to see are Kim's Nomological Sufficiency Conception of Causation and the Causal Closure Principle: The first says that A causes B only if A is nomologically sufficient for B (Kim 1993b: 351); the second says that any physical event that has a cause at t has a complete physical cause at t (Kim 1989: 43). If there is downward causation from  $M_1$  to  $P_2$ , then, by the Nomological Sufficiency Conception,  $M_1$  is nomologically sufficient for  $P_2$ . However, by the Causal Closure Principle, if  $M_1$  causes  $P_2$ , then  $P_2$  has a complete physical cause  $P_1$ . Now  $P_1$  is nomologically sufficient for

$P_2$ , according to the Nomological Sufficiency Conception of Causation.

On the assumption that  $M_1$  is not identical to  $P_1$ , we have two distinct nomologically sufficient conditions for  $P_2$ , namely,  $M_1$  and  $P_1$ . However the instance of  $M_1$  is there because, according to Kim, it has its own physical realization base,  $P_1$ , which is sufficient, non-causally, for  $M_1$ . The Physical Realization Thesis claims exactly this:

[The Physical Realization Thesis]: A mental property is instantiated only if it is realized by a physical property. If  $P$  realized  $M$ , then  $P$  is nomologically sufficient<sup>28</sup> for  $M$ , and  $M$  supervenes on  $P$ . (Kim 1993b: 347)

Since  $P_1$ ,  $M_1$ 's physical realization base, is non-causally sufficient for  $M_1$ , it follows that  $P_1$  is sufficient for  $P_2$ .

Now we face a serious difficulty, the problem of Causal/Explanatory Exclusion:

[The Principle of Causal/Explanatory Exclusion]: There is no more than one complete and independent cause (or causal explanation) of any event. (Kim 1989a: 250)

---

<sup>28</sup> Here it should not be, of course, causally sufficient.

Kim claims that if  $M_1$  and  $P_1$  are distinct nomologically sufficient conditions for  $P_2$  and  $P_1$  is nomologically but non-causally sufficient for  $M_1$ , then  $P_1$  is the only genuine cause of  $P_2$ . Kim says, "The more basic causal relation obtains between the two physical properties,  $P_1$  and  $P_2$ , and  $M_1$ 's causation of  $M_2$  is ultimately grounded in the causal relation between their respective physical realization bases" (1993b: 353). He further says that:

All these considerations, I want to suggest, point to something like the following as the natural picture for the layered physicalist world: all causal relations are implemented at the physical level, and the causal relations we impute to higher-level processes are derivative from and grounded in the fundamental nomic processes at the physical level. ... [I]f, as the supervenience thesis claims, all the facts are determined by physical facts, then all causal relations involving mental events must be determined by physical facts (presumably including facts about physical causation). (Kim 1993b: 355)

From this consideration Kim elicits the problematic principle, the Causal Inheritance Principle:

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



This principle, which guarantees that no higher-level property-instance confers on its bearer any new causal powers, however, is the very principle that opens the door to an accusation of epiphenomenalism. If  $P_1$  is the only genuine cause of  $P_2$ , and  $P_1$  is not identical to  $M_1$ , then  $M_1$  does not cause  $P_2$ . If  $M_1$  does not cause  $P_2$ , then  $M_1$  does not cause  $M_2$  because of the Causal Realization Principle. Therefore,  $M_1$  does not cause  $M_2$ , and so,  $M_1$  is epiphenomenal.

Kim thinks that this, taking  $P_1$  as the cause of  $P_2$  and treating  $M_1$  as epiphenomenal, is a persuasive picture. Faced with the question, "Is there any reason for invoking  $M_1$  as a cause of  $P_2$  at all, given  $P_1$  is sufficient physical cause of  $P_2$ ?" Kim's answer is clear: no causal powers over and beyond those of  $P_1$  are left for  $M_1$ . The whole point is that if nonreductive physicalists accept downward causation by irreducible mental properties, they should accept a problematic principle, the Causal Inheritance Principle. And Kim claims that the exclusion problem raised from the persuasive picture of downward causation is the problem that nonreductive physicalism cannot deal with.

### 3.3 AGAINST THE EXCLUSION ARGUMENT

I, like every philosopher, do not believe that we can make sense of the world without supposing that the mental properties are causally efficacious. One of my strategies in dealing with the problem raised by the Exclusion Argument is to argue against the Causal Inheritance Principle (CIP), by showing the causal relations between mental properties  $M_1$  and  $M_2$  do not depend on causal relations between the properties that realize them.<sup>29</sup> Before we turn to CIP, against which I will argue in the next section, let me suggest the claim that the causal relations between mental properties  $M_1$  and  $M_2$  do not depend on causal relations between the properties that realize them.

Suppose that we want to explain James' promising to his mother to go to church, and that the putative explanation is that James wanted to please his mother, and believed that James would do so by promising to his mother to go to church. The explanatory connection is between James' belief/desire complex and James' promise.

---

<sup>29</sup>The claim that the causal relations between mental properties do not depend on causal relations between the properties that realize them is also developed in Baker (2001).

Suppose that as a result of James' promising to his mother to go to church, his mother was happy. Suppose that James' promising to his mother to go to church,  $M_1$ , was realized by microproperties  $P_1$  and that his mother's being happy,  $M_2$ , was realized by microproperties  $P_2$ . The mother's being happy is causally explained by James' promising to his mother to go to church. But it by no means follows that  $P_1$  causally explains  $P_2$ . The assumption that  $P_1$  must causally explain  $P_2$  is an artifact of a reductive picture.

If we focus on mental properties,  $M_1$  and  $M_2$ , that  $P_1$  and  $P_2$  realize, then it is apparent that the causal relations between mental properties do not depend on causal relations between microproperties that realize them. Which microproperties realized James' promising to his mother depends on how the promise was made (e.g., by making a phone call, or by writing a letter, etc.). But the effect of that promise - his mother's being happy - is indifferent to how the promise was made (by making a phone call, by writing a letter) and thus indifferent to which microproperties realized the promise. James' promise would have had the same effect no matter which microproperties realized it.

Counterfactual conditions play a large role in our understanding of causation. The truth of a relevant counterfactual is a *typical* indication of causation. It is typical because not all counterfactuals are causal (Kim 1993c: 205-207). However, the truth of a relevant counterfactual is clearly a necessary condition for causation. If James had not wanted to please his mother, nor believed that by promising he would please her, James would not have promised to his mother to go to church (unless James had some other reason). There need be no relevant counterfactual, between the properties that realized James' belief/desire complex and the properties that realized the promise. Let me explain this a little further. An instantiation, by James, of the property  $M_1$  (e.g., James' promising to his mother to go to church) causes an instantiation, by his mother, of the non-mental property,  $P_2$ . It happens because  $M_1$  causes  $M_2$ . The relevant counterfactual should be: if James had not promised to his mother, there would have been no instantiation of  $P_2$ .

By contrast, there may be no relevant counterfactuals between the non-intentional properties that happened to constitute James' promise and the

nonintentional properties  $P_2$ . To see this, suppose that James' promise was constituted by writing a letter to his mother, which, in turn, was constituted by a left-to-right motion of James' right hand. Now it is clearly wrong to say that if James' hand had not moved left-to-right in the circumstances, then there would have been no instantiation of  $P_2$ . The relevant circumstances are the circumstances in which you were intending to make a promise. In those circumstances, even though James' right hand had not moved left-to-right, James would have made the promise some other way - e.g., by making a phone call to his mother and saying he is going to church. The only relevance of his hand's moving left to right was that the motion constituted James' promise.

The effect of the promise is James' mother's being happy, and James' mother's being happy is realized by  $P_2$ . What has the effects on her reaction is the promise, not what realizes the promise. The properties whose instantiations realize the promise are typically irrelevant to the mental effects of the promise. So we can account for the causal relations of James' belief/desire complex causing James' promising. We can also account for the causal relations of James' promising

causing his mother's being happy. In addition, we can account for causal relations between intentional properties and their non-intentional effects - James' promising to go to church which caused his mother's being happy, caused instantiation of the nonintentional properties,  $P_2$ , that realized James' mother being happy. But if mental property,  $M_1$ , causes mental property,  $M_2$ , and  $M_1$  is realized by non-intentional properties  $P_1$  and  $M_2$  is realized by non-intentional properties  $P_2$ , it does not follow that  $P_1$  causes  $P_2$ .

This provides a conclusive reason to reject CIP, the heart of the Exclusion Argument. The Causal Inheritance Principle is false, because the causal powers of particular instantiations of mental properties are not inherited from the non-intentional properties that realize them. As we assumed,  $M_1$  causes  $M_2$  and  $M_1$  and  $M_2$  are realized by physical states,  $P_1$  and  $P_2$ , respectively. According to CIP, the causal powers of the instance of  $M_1$  are identical with the causal powers of  $P_1$ . Then the relations between  $M_1$  and  $M_2$  *do* depend on causal relations between  $P_1$  and  $P_2$  that realize  $M_1$  and  $M_2$ . However we saw that the causal relations between mental properties do

not depend on causal relations between the properties that realize them. Therefore, CIP is wrong.

We should not accept CIP because which non-mental properties realized  $M_1$ , depends on *how* it was made. The effect of  $M_1$ , however, is not affected by *how* it was made. Whichever way  $M_1$  is realized, it has the same effect,  $M_2$ . This is the subject of the next section.

### **3.4 AN ARGUMENT AGAINST CIP: CONSIDERATION FROM CONTENT EXTERNALISM<sup>30</sup>**

CIP says that a mental property, realized in virtue of a physical realization base, has no new causal powers beyond the causal powers of physical base. It claims that the causal powers of higher-order properties can be explained through the implementing mechanism. Kim's rationale to elicit CIP is the following consideration: each psychological explanation requires some physical implementing mechanism. Therefore, the psychological properties inherit their causal powers merely from the

---

<sup>30</sup> What is strange, though, is that Kim has never tried to solve the problem raised from content externalism: it is strange because his Exclusion Argument cannot be used without solving the problem raised by content externalism. As his article "Psycho-physical Supervenience" (1982) reveals, he also seems to have strong sympathies for some notion of narrow content. This seems to be confirmed in that Kim tries to keep type-identity theory by using so-called local reduction. See his (1998).

physical properties of the implementing mechanism. Unlike the physical explanations considering only the lower-level properties, however, we need to consider specific social and/or historical environments in order to explain an action. Mental explanation has a much broader context than physical natural explanation.

Rejecting CIP seems to imply that the causal powers of mental properties somehow magically emerge at a higher-level and there is no accounting of the new causal powers of mental properties in terms of lower-level properties and their causal powers and nomic connections. If we follow Kim's distinction between micro-based higher-level properties and higher-order properties, we can see that the causal powers of micro-based properties emerge from their micro-structure, which means the seeming new causal powers are not magical. This is the reason that Kim thinks that CIP does not apply to micro-based macro properties.<sup>31</sup> However, unlike Kim, I don't think we should see the new causal powers of higher-level properties as emerging magically, either. The reason, I think, that CIP does not work even in the case of higher-

---

<sup>31</sup> "[Micro-based properties] need not be, and are not likely to be, identical with the causal powers of these constituent properties and relations" (Kim 1998: 117).



order properties, is because social and/or physical environments that are constitutive for mental explanations involve essentially a mental dimension. In this case the supervenience base has wider base than just implementing physical states. Causal mechanisms considering only the lower-order properties in no way reflect this wider base. The implementing mechanism is not able to describe the causal powers of higher-order properties resulting from the interaction with social and/or physical environments.

In this section I will argue against CIP by using the lesson learned from content externalism. Before doing that, however, let me draw your attention to the difference between mental explanation and naturalistic explanation with regard to the why- and how-questions.<sup>32</sup> At a general level, we can characterize explanations as answers to certain kinds of questions. For example, in science and various mundane contexts, mechanistic explanations are taken to answer the questions, "why some events happen," as well as "how some events come into existence." Since it may appear initially plausible that why-questions about actions require causal answers,

---

<sup>32</sup> This point will be examined and clarified in Chapter Six.

citing a mechanism is, therefore, often taken to give a causal explanation; they exist because they are caused by other events. In the case that the why-question is interpreted as a request for a mechanism which we may or may not be able to provide, any response (explanations) to the why-question also provides information (explanations) that can adequately answer the how-question. In mental explanation, however, it is not the case that we expect the same pattern of interchangeability between why- and how-questions.<sup>33</sup>

Kim, unlike me, thinks that in mental causation mental explanations are answers to why-questions in the sense that they are using only the implementing, lower-level physical bases that are grounded in objective relations. However, mental explanations in mental causation, which answer our why-questions, do not seem to describe objective relations. I insist that naturalistic explanations describe objective relations but do not answer why-questions. Therefore they neither adduce causal information nor provide the explanatory answers. This is the lesson we have learned from content

---

<sup>33</sup> In the case of mental causation I take mental explanations as answers to why-questions and physical explanations as ones to how-questions.

externalism. Mental explanations are interested in explaining phenomena interacting with a wider social and natural environment. Mental explanations are not in competition with explanations introducing lower-level implementing mechanisms since mental explanations take as their primary subject of explanation an action in so far as the action is interacting with a certain environment and is directed towards that environment. A physical implementing mechanism is not able to explain this interaction with environments. The new causal powers have not magically emerged; they arise from interaction with the environment to which we, as agents, are related. It is thus hardly surprising that the causal power of higher-level properties, interacting with the environment, cannot be described on the physical level.

Widely individuated content has different causal powers from those of implementing physical states. Even if mental explanations require certain lower-level physical implementing mechanisms, this does not show that mental properties do not have the causal powers beyond those of the lower-level physical properties. I argue from this consideration that NCC is the result of confusing a purely naturalistic explanation with mental

explanation. The rejection of CIP entails the rejection of NCC. As we saw, the causal relations between mental properties do not depend on causal relations between microproperties that realize them. Then, NCC, which says that all events related as cause and effect fall under strict law, is false. The causal pattern at mental levels, which can occur only in certain circumstances, is not governed by the causal patterns at the lower levels since they cannot be explained by the non-intentional realizing properties which do not consider matters interacting with the context or circumstances. I will defer further discussion of this issue until the final chapter, however, because the issue is closely related to the issue of intentional actions, which is the subject of the second part of this work.

### **3.5 THE EXCLUSION ARGUMENT: CONSIDERATIONS FROM EXPLANATORY PRACTICE**

It is possible that the events quantified over in the categories of the mental, the social, or the biological will turn out to be the very same events quantified over in one, very special and extraordinary,

explanatory theory; however, it isn't likely. We can make this claim more secure by noting the different methodological commitments involved in the sciences; and, more significantly, how even within a science, convergence on one ontology is difficult to come by. Many disciplines possess methodologies and explanatory devices that researchers in other disciplines find highly suspect. Such difference makes it difficult to see how the objects of such diverse sciences could be identical.

As Dupre (1993) has argued, convergence on a common ontology within a discipline cannot be assumed, even when the theoretical terminology, and the ontological commitment that follows from the employment of such concepts, appears to be unified. Dupre points out that " ... in some contexts species are treated as individuals, in others as kinds" (Dupre 1993: 42). What is particularly important about Dupre's work is that it reveals how the ontology of one theory can be quite different from that of another theory in which the theoretical terminology is shared. The species concept may pick out an individual or a kind, depending on the explanatory context. It should come as no surprise that sciences that differ in methodology and in explanatory

goals or concerns should be committed to distinct ontologies. After all, the methods and explanatory goals have been formulated and developed in order to best suit the subject matter under scrutiny. Differences in methodology and explanatory concern are likely to reflect differences in the ontology that these methods and concerns have been brought to bear upon.

There is some form of dependence between the mental and the physical. Global supervenience is such a dependence relation. Still we don't know how those events are related. This is the reason why we need to pay attention to our practice and explanatory strategies. I don't think the demand for strict laws is, as NCC claims, essential to causal relations. However, the motivation for the demand for laws in causal relations stems at least in part from the fact that the laws cited in explanation are the laws that subsume events in naturalistic causal relations. Many accepted psychological causal explanations, however, like many explanations in general, do not cite laws. There is no reason to accept the claim that psychological causal explanations cite causally relevant (or causal/explanatory) properties, but the only causally

efficacious properties (or "genuinely" or "robustly" causal properties) are those of physics (or those reducible to physics).

I favor the view of intentional causation in which true intentional causal explanations are grounded in causal relations in which mental particulars play causal roles in virtue of their intentional properties. We do have a great deal of evidence for this: what we think affects what we do. We have an overwhelming amount of both scientific and non-scientific evidence about the causal relations between belief/desire complexes and actions. However, we have no evidence at all about the causal relations between the instantiations of the non-intentional properties that realize belief/desire complexes and the instantiations of the non-intentional properties that realize actions. Our conviction that what we think affects what we do is more secure than any metaphysical argument against it.

In Chapter Four I will respond to some challenges to this conviction, by using Baker and Burge's proposal to think about the causal efficacy of specific properties in the context of established scientific and commonsensical explanatory practices. Burge's point regarding causation

can be understood as asserting that we must formulate our metaphysics of causation against our background knowledge of actual causal/explanatory practice. We should not approach the nature of actual causal/explanatory practice with a priori assumptions regarding causation. We shall learn about the nature of causation by examining how causation features in our explanatory commitments. If we have informative and fruitful mentalistic explanation, then we have every reason to believe that mental events exist and interact.



## CHAPTER 4

### COMMONSENSE PSYCHOLOGY AS AN EXPLANATORY PRACTICE

The problem of mental causation appeared when we wanted to confer some kind of primacy to the physical without abandoning the autonomy of the mental. I argued in the previous chapters that a particular unanalyzed assumption, NCC, is responsible for a philosophical impasse. Modifying our conception of causation would, I suggest, leave us with a means of reconciling our various intuitions concerning the nature of the mental, and give us an adequate account of the causal relevance of psychological and other supervenient properties.

Chapter One discussed a problem of mental causation by exploring Donald Davidson's AM. We saw that AM is committed to the epiphenomenalism of the mental. I claimed that NCC is not something that we can tolerate. In Chapter Two I dealt with a tension that arises from content externalism. I examined debates, one between Burge and Fodor, the other between Burge and Davidson. From the first debate I argued that there is no *a priori* reason why the so-called "wide" contents do not or cannot play causal roles in psychological explanations of

behavior, and showed how they might do so by noting that wide contents are among the properties we ordinarily cite to explain our behavior. The result I elicited from both debates was that we have good reasons for rejecting NCC. Fodor basically argued that individuals cannot have different causal powers without different brain states. Davidson's NCC, according to Burge, is a more imaginative version than Fodor's claim, but claims the same point as Fodor: "physiological processes are where the "real" causation in psychology goes on" (Burge 1989a: 306).

The Exclusion Argument is designed to show that nonreductive conceptions of the mental face the serious problem of producing an account of mental causation which does not render the mental epiphenomenal. In Chapter Three I showed that the solution to the Exclusion Argument was reached by rejecting the Causal Inheritance Principle (CIP). I argued that the rejection of CIP actually implies that NCC is in fact wrong.

In the present chapter a new conception of causation starts to emerge as a result of rejecting NCC. In Chapters Five and Six this conception will be discussed with regard to intentional actions. In this chapter I will first explain commonsense psychology (hereafter CP),

and then argue against the claim that CP is a kind of a scientific theory.<sup>34</sup> The alternative to regarding CP as a scientific theory is to regard it as a practice. Secondly, I will argue that our explanatory practice should guide our ontological commitment. And, finally, I will defend my position against what I see to be a number of serious challenges.

The primacy of explanatory practice over the ontological commitment reverses the usual account in which causal explanations count as causal if they are grounded in causal relations. However, explanations come first, such that an explanation is causal if we accept it as such.<sup>35</sup> By reinterpreting the notion of causation we regain the causal efficacy of the mental. The problem raised by the Exclusion Argument, I claim, takes a wrong point of departure when it begins with a metaphysical notion of causation instead of grounding the notion of causation on our explanatory practices.

---

<sup>34</sup> I will use the word broadly in a sense that something is called a scientific theory when it can be falsified by a mature science.

<sup>35</sup> This position is similar to the one that I will explain with regard to intentional actions. The usual account of intentional actions takes an action as intentional if it is grounded in reason explanation. I will reverse the account, and that is the main issue of Chapters Five and Six.

This strategy has been defended by Baker (1993), Burge (1993), and van Gulick (1993). The mental, they argue, is causally relevant or efficacious only insofar as it figures in successful explanations. Baker, for example, explicitly rejects the metaphysical picture of physicalism, which "subordinates explanation to causation, where causation, in turn, is conceived as an 'objective relation' in nature" (1993: 93). In her terms, "causation becomes an explanatory concept" (1993:93): causes are the sorts of things that are cited in explanations of events. She would insist that the success of our explanatory practices is enough to ensure that any metaphysical assumptions that lead to an epiphenomenalist conclusion must be wrong. We have more confidence in the success of mentalistic explanation, typical commonsense psychological statements that refer to mental states as causes of behavior, than we do in the basic tenet of physicalism according to which causation involves physical events and properties as causes.

There are some serious challenges that this conception appears to face. (1) It has not always been accepted that rationalizing explanations are causal explanations, so common practice does not obviously

assume causal relevance (Kim 1995; 1998). Many philosophers, such as Melden (1961) and Kenny (1963), between the late 1950s and early 1960s, influenced by the later work of Ludwig Wittgenstein, rejected the view that the relation between reasons and actions is a causal relation. The assumption, so the objection goes, that common explanatory practice assumes causal relevance may simply not be true, and is certainly not justified without additional argument. (2) Even if explanatory practice assumes that the mind is causally relevant, this fact does not explain how it is possible for the mind to be causally relevant. It does not provide an answer to the more philosophically important question of how mental causation may occur. (3) Explanatory practice is defeasible, and the Exclusion Argument may provide reason to defeat it. I will examine these challenges in turn, and reject them. In section 4.1 I claim that CP is not a kind of a scientific theory but a practice. After that I will deal with each of the three challenges.

#### **4.1 COMMONSENSE PSYCHOLOGY (CP) NOT AS A SCIENTIFIC THEORY BUT AS A PRACTICE**

CP concerns the ordinary psychology of beliefs, desires, and emotions for accounting for each other and ourselves. It tries to explain behavior by reference to certain types of mental states, mental states with propositional content such as beliefs and desires. It is a tool for predicting and explaining behavior. For example, CP asserts that, if someone desires that  $p$ , and believes that  $\Phi$ -ing will satisfy that desire, then, *ceteris paribus*, that person will  $\Phi$ .

Most critics and defenders of CP endorse the materialist assumption that intentional psychological phenomena - if they exist at all - are incarnated in the human brain. Most critics and defenders of CP also assume that CP explanations will not reduce to neurophysiological explanations. Critics of CP see this "failure" as a reason for rejecting the postulated ontology of CP, whereas defenders of CP see it as a reason for maintaining the autonomy of commonsense psychological explanation.<sup>36</sup>

---

<sup>36</sup> The situation is similar to the problem of mental causation in that we want to confer some kind of primacy to the physical without abandoning the autonomy of the mental. A vast majority of contemporary views want it both ways: the physical is primary but the mental is real and distinct from it.

There are two basic ways in which CP may be approached by philosophers. First, it is considered as a sort of proto-science: CP is developing a scientific form of explanation. They consider it as a theory about the internal causes of our actions, potentially in competition with scientific ways of explaining behavior, and vulnerable to being shown false. Those who construe CP as a kind of a proto-science emphasize a metaphysical notion such that what happens is subject to integration into the physical sciences. The second way to see CP is to take it as a different sort of activity, not as scientific or proto-scientific theorizing. Philosophers who take this position see CP as an autonomous *explanatory practice* (Baker 1999), not in competition with science nor threatened by it. As Mele points out, any adequate philosophical analysis of intentional action should be anchored by commonsense judgments about particular cases (Mele 2001).

I am attacking the first sort of view, and defending the second. I argue for the truth and legitimacy of commonsense, propositional-attitude-based explanations of behavior, but not on the grounds that a naturalistic explication or reduction of propositional attitudes is

likely to be forthcoming. Rather, I argue that there is no good reason not to accept the legitimacy of an autonomous rational psychology construed as explanatory practice.

The following definition of CP as a practice will work for our purpose:

Commonsense psychology [CP] is a practice iff groups of people engage in the activity of describing, explaining and predicting human thought and action in terms of propositional attitudes like belief, desire and intention. (Baker 1999: 4)

Some of our practices involve giving causal explanations. I take CP as a causally-explanatory practice, a practice governed by rules or conventions that people engage in for a common purpose. Because of the success and wide acceptance of commonsense psychology, debates here instead focus on the criteria for specifically causal explanations and whether psychological explanations meet these criteria.

Baker's solution for the problem of mental causation is to rethink CP and the notion of causation that generates the problem of accounting for the causal efficacy of non-physical properties. She says, "Systematic explanatory success, in either science or



everyday life stands in no need of metaphysical underpinning" (1993: 94). The idea is to put aside worries about the causal efficacy of non-physical properties by uprooting the assumption that only physical properties can have a causal impact on the physical world. Geological, biological, meteorological, psychological properties, and so on, do figure into explanations that seem to rely on causal relations between them and physical events and properties. The idea that the most basic physical properties might somehow "gobble up" all causal efficacy of the macro-level, that they provide the "complete cause" of physical effects, seems to undermine common sense and scientific practice. Hence, according to Baker, we should not think that there is any problem with mental causation in particular, because our explanatory practices provide stronger confirmation of its reality than the claim that all causation involves physical properties. For as she points out, we don't know much about the most basic physical properties of the world. However, we know a lot more about macro-properties and their relations. Our insistence that the bottom level provides all causality makes it seem as if we have betrayed commonsense in favor of a rather obscure

commitment to causality as an objective relation - one distinct from our explanatory practices and epistemology.

Burge (1993) accepts that mental content does not supervene on the physical but does not see this failure as impacting the problem of mental causation. Burge relies on explanatory practice and our ordinary notion of causal powers to allow for mental causation. He argues that common explanatory practice picks out some regularities as causal, and since this explanatory practice assumes mental-to-mental causal relevance or mental-to-physical causal relevance, mental properties are causally relevant. Burge realizes that relying on regularities alone fails to distinguish epiphenomenal from causally relevant properties, but requires instead that common explanatory practice be our guide in picking out the causally relevant properties.

Burge's point regarding causation can be understood as asserting that we must formulate our metaphysics of causation against our background knowledge of actual causal-explanatory practice. We should not approach the nature of actual causal-explanatory practice with a *priori* assumptions regarding causation. Furthermore, we should not pronounce (metaphysical) judgment on the

status of explanations from disciplines such as psychology with such *a priori* assumptions. At the very least, we should not attempt revisionary theories and practices regarding such causal explanations and their *prima facie* ontological commitment. Rather, we shall learn about the nature of causation by examining how causation features in our explanatory commitments. If we have informative and fruitful mentalistic explanation<sup>37</sup>, then we have every reason to believe that mental events exist and causally interact. Again, our causal-explanatory practice and the natural ontological commitment stemming from such practice should determine our metaphysical commitments.

There are objections that appeals to explanatory practice alone are insufficient in solving the problem of mental causation. Thus Kim (1995) says, the assumption that common explanatory practice assumes causal relevance may simply not be true, and is certainly not justified without additional argument. Some substantive theory of mental causation that takes into account the Exclusion Argument is necessary to solve this problem of mental

---

<sup>37</sup> By mentalistic explanation I mean typical folk psychological statements that refer to mental states as causes of behavior.

causation. In the following three sections I will deal with three objections directed toward the idea that we should appeal to explanatory practice.

#### **4.2 RESPONSE TO THE FIRST CHALLENGE**

In this section I deal with the first challenge, saying that it has not always been accepted that rationalizing explanations are causal explanations, so common practice does not obviously assume causal relevance. As Kim points out (Kim 1998: 63), the assumption that psychological explanation, like much scientific explanation, is causal in nature was itself a source of heated debate in philosophy during the 1960's. For instance, philosophers thought that rationalizing explanations were not a variety of causal explanation at all. One cannot simply assume that the common practice of intentional and reason explanations is causal. A central thesis of many neo-Wittgensteinian accounts was that folk psychological references to intentional psychological states are not causally explanatory. It was Donald Davidson who managed to convince a majority of philosophers that reason-giving explanations are a form

of causal explanation (Davidson 1963). But that argument involved a theory of causation, events, and explanation. This shows that taking explanations as our starting point does, itself, require various metaphysical commitments. Why should we assume, with Baker, then, that such folk psychological explanations are causal? If we do, it seems we have already presupposed a lot of metaphysics. The problem of mental causation can be seen as the attempt to sort out those assumptions to help understand just what sort of "metaphysical underpinning" we have available.

The defenders of explanatory primacy might have an answer to Kim's point, though. It may be said that the choice of making causation dependent on explanatory practices is itself a metaphysical choice. Kim does not have to be budged by this, because Kim and others can argue that what the defenders of explanatory primacy are doing is giving up a view according to which there has to be an objective relation grounding the relation between the explanandum and the explanans. And if they subordinate causation to explanatory practice, there will be a danger that we would do the same with other dependence relations such as supervenience. If what is real at least in part depends on what is involved in

causal or dependence relations, and causation and other dependence relations are dependent on explanatory practices, we may not be able to avoid the anti-realist consequence that Baker wants to avoid.<sup>38</sup>

I doubt that the objection is successful. As we saw in Chapter Three, for the response to the Exclusion Argument I used an argument that has lots of metaphysical implication. The response does not have any anti-realist flavor, however. I just rejected CIP and paid attention to the implications of content externalism. Therefore it is not legitimate to say that the emphasis on explanatory practice has no metaphysical basis. I provided an argument against the Exclusion Argument over metaphysical commitment and I chose explanatory practice based on this argument. My choice is the result of serious metaphysical considerations. It is not the case that explanatory practice is a groundless idea without any metaphysical implication. I have not taken this view for granted.

Let us look at the following causal explanation:  
James promises his mother to go to church because of

---

<sup>38</sup> Baker says the following: "Although my proposal has a strong pragmatic cast, it is by no means an anti-realist suggestion. I am not equating what is real with what is needed for explanations and predictions" (Baker 1993: 95).

James' wanting to please his mother. The promising to his mother to go to church is the kind of thing that we want to explain; in other words, we want to know why James promised his mother to go to church. The very existence of the *explanandum*, however depends on rules, practices, or conventions. In this example for instance, apart from the religious practice of going to church and the practice of performing a promise, there would be no such phenomenon as somebody's promising to his mother to go to church. In the absence of rules, practices and conventions, what we want to explain would disappear. Therefore, a putative explanation of any of these things in terms of, say, physical motions, without reference to rules, practices and conventions, is no explanation of what we set out to explain at all - namely why James promised to his mother to go to the church. I take this as a lesson learned from content externalism, which means I have paid enough attention, metaphysically speaking.

#### **4.3 RESPONSE TO THE SECOND CHALLENGE**

The second objection is the claim that those who favor explanatory practice over metaphysics do not

provide an answer to the philosophically more important question of how mental causation may occur. As Kim points out, even if explanatory practice assumes that the mind is causally relevant, this fact does not explain how it is possible for the mind to be causally relevant. The problem of mental causation is not that we do not think the mind is causally relevant but that we do not have a metaphysical picture of the mind and the world that allows for the mind to be causally relevant. The question, then, is not so much whether the mind is causally relevant, but rather how it is possible for the mind to be causally relevant. And Burge's appeal to common practice does not answer this question. The appeal to common practice misplaces the origin of the problem of mental causation. Unless we are ready to discard metaphysical questions as significant ones, we have to recognize that there is a conflict between different assumptions we make and that the problem will not go away if we don't give up or reformulate some of these assumptions. Kim says:

The issue is not metaphysics versus explanatory practice, as Burge would have it, nor metaphysics versus epistemology, as Baker would have it ... The issue is *how to make our metaphysics consistent with mental causation*, and the choice we need to make is



between various *metaphysical alternatives*, not between some recondite metaphysical principle on the one hand and some cherished epistemological practice or principle on the other. (Kim 1998: 62)

Kim seems to claim in the above passage that we need to provide an account that supports both our commitment to mental causation and the metaphysics behind it.

Why do we expect a causal story that makes reference only to neurophysiological phenomena? And why would the success of neurophysiology provide good reason to take it seriously as a domain of legitimate causal explanation? Now, the objectors insist further that if neurophysiological explanations are distinct from intentional explanations, we are left with a mystery: the mystery of how they relate to one another. In order to solve the mystery an eliminativist, for example, argues that since all the causal linkages here are purely neurophysiological in nature, any alleged "mental causes" are unnecessary and hence should be sliced off with Ockham's razor.

However, the requirement of our having a bottom-level or neurophysiological process seems to follow from NCC. However, as we saw in the previous chapters, the requirement of there being an ontological grounding for

the intentional phenomena including mental causation is just the myth of physicalism.

The problem, if we are accepting NCC, can be expressed in the following way: how are causation and causal explanation related? The distinction between causation and causal explanation is that, while one relation holds between natural entities whether or not we exist, the other is a conceptual relation between linguistic entities (or perhaps propositions) when we find that the one illuminates the other.<sup>39</sup> The most widely accepted view is that the former provides the ontological grounding for the latter; a true causal explanation counts as causal because there is, behind it, an instance of causation (Kim 1989a:254-260). In the best case, the causal relation that grounds a causal explanation holds between events in virtue of those properties denoted by the predicates that play the appropriate roles in the explanation.<sup>40</sup> However, as the extensional view of causation shows, the features in virtue of which a

---

<sup>39</sup> The extensional view of causation relies on a distinction between descriptions that can appear in singular causal claims and those that, in addition, denote causally efficacious properties of tokens. This is one way of expressing a certain relation between causation and causal explanation.

<sup>40</sup> In Chapter One we saw that Davidson argued against this view. I showed that his argument was not successful.

certain causal relation holds need not be mentioned. A causal explanation can be ontologically grounded in a causal relation even if it does not specify its ontological ground by referring to the property of the object that is causally efficacious in that relation. Now the issue is whether good causal explanations require laws. The view that they do, a position I have attacked, dovetails with the nomological account of causation: the causally related events stand in a causal relation in virtue of the fact that they can be subsumed under a law.

However, if we reject NCC, we don't need to worry about finding some ontological ground relating intentional explanations to physical explanations. The central point is supposed to be that the singular causal statements we invoke in action explanations are not in need of any appeal to regularity or law, but are themselves legitimate. The motivation for the demand for laws in action explanations stems at least in part from the fact that the laws cited in explanations are the laws that subsume events in natural causal relations. By rejecting the idea that causal explanation is causal because it is grounded in natural causal relations, the motivation for requiring laws in explanations disappears.

In addition to this fact, many accepted psychological causal explanations, like many explanations in general, do not cite laws. We have another strong reason to reject NCC. Baker writes:

For example, when Jill returns to the bookstore to retrieve her keys, what she thinks is that she left her keys on the counter and that she wants them back. What she thinks affects what she does in virtue of the following explanatory fact: if she hadn't thought that she had left her keys, then, other things being equal, she wouldn't have returned to the bookstore. (1993: 93)

As we have seen in Chapter Three, the truth of a relevant counterfactual is a *typical* indication of causation, typical because not all counterfactuals are causal; however, the truth of a relevant counterfactual is clearly a necessary condition for causation.

Now unless Kim and others are ready to discard a physicalistic picture as the only genuine one, they are not able to see where they are wrong. We already saw that the causal relations between mental properties do not depend on causal relations between the properties that realize them. I argued for this not as metaphysics versus epistemological practice or principle. I argued that the properties whose instantiations realize the mental are typically irrelevant to the effects of the mental. There

is no need, then, to expect a causal story that makes reference only to neurophysiological phenomena. There is no mystery between intentional explanations and neurophysiological explanations of how they are related to one another. In some ways they are related, as in some form of mind-body supervenience, and we may not know the exact nature of the relation. However the ignorance is not a mystery.

#### **4.4 RESPONSE TO THE THIRD CHALLENGE**

Now let us look at the last objection. It claims that explanatory practice is defeasible, and the Exclusion Argument may provide reason to defeat it. Our common practice may be mistaken. In this case, we may mistakenly attribute causal relevance to mental properties. Scientific considerations have often overcome common practice. Perhaps the case of mental causation is another case in which scientific considerations, suitably informed by philosophy, should overcome our common practice.

I undertake my defense of CP as an autonomous explanatory practice by first undermining the opposing view, Eliminative Materialism (hereafter EM). EM is the view that CP is a theory, which is in competition with scientific theories, and likely to be proven false. EM does not consider CP as a viable theory and should therefore be rejected. According to Patricia Churchland, by EM, she means:

(1) that folk psychology is a theory; (2) that it is a theory whose inadequacies entail that it must eventually be substantially revised or replaced outright (hence "eliminative"); and (3) that what will ultimately replace folk psychology will be the conceptual framework of a matured neuroscience (hence "materialism"). (1986: 396)

Taken as applying to CP instead of to folk psychology,<sup>41</sup> Churchland's definition of EM is highly questionable.

First, is CP a theory? It seems that CP is used to

---

<sup>41</sup> My concern on the usage of the term "folk psychology (FP)" is that there are at least two ways in which the term might be used. FP might be used to mean that pre-scientific psychological theory, implicitly held and used in everyday life, by "the folk," namely ordinary, unsophisticated persons. Such a FP presumably would include pre-scientific speculations and preconceptions regarding the nature of all sorts of psychological phenomena: mental illness, sleep and dreams, motivation, problem-solving, perception, and so on. Psychologists might tend to use the term FP in this way and to take it as an empirical matter.

On the other hand, FP might be used as philosophers tend to use the term, to refer to the practice of predicting and explaining behavior by reference to propositional attitudes. However, if FP is just whatever the folk think about psychology, then CP is only one aspect of FP. Since I am primarily concerned with the legitimacy of CP, and not with the status of whatever else has been called FP, I will henceforth avoid the use of the confusing term FP and use the term CP instead.

describe rational capacities, which can function as an explanatory and predictive system by subsuming individual actions under generalizations involving the described capacities or properties of rational systems. For example, CP described persons as believing that *p*, perceiving that *p*, wanting that *p*, intending that *p*, and so on.

Individual behavioral events can be explained by subsuming them under generalizations involving these properties, as in the following example: Users of CP implicitly know some such generalization as if X believes that there is poison in the glass in front of him, then, *ceteris paribus*, he will not drink the contents of the glass. We may explain why X did not drink his wine on a certain occasion by reporting that X had a certain propositional attitude property: "He believed that there was poison in it."

Certainly, there are disanalogies between CP and the classical sort of empirical theory that postulates unobservables, and articulates generalizations regarding the behavior of those unobservables, in order to explain observed data. CP implies that rational beings possess propositional attitude states (properties). When we utilize CP to predict and explain the behavior of others,

what we are doing is projecting onto others an explanatory system experienced firsthand in our own case. We know that our own belief and desire states or properties explain our behavior, and we project ourselves into other persons' situations, asking ourselves what we would believe and desire, and what we would do, if we were in that situation.<sup>42</sup>

This sort of projective practice, based upon first-person experience, does not resemble classical theoretical explanation. It does not involve unobservable entities, and the generalizations of CP bear little resemblance to the generalizations of a typical empirical theory. When someone suggests what the generalizations of CP might be, the suggested candidates are always instances of principles of practical rationality, such as if X believes that *p* only if *q*, and if X desires that *p*, then, *ceteris paribus*, X will try to bring it about that *q*. The generalization mentioned above, involving the poisoned wine, may be seen as an instance of such a principle of practical rationality: if X desires to live,

---

<sup>42</sup> The suggestion that CP is "projective" in this sense has been made by Robert Gordon (1986). Stephen Stich (1983: ch.5) has also made remarks to this effect. The so-called simulation theory has been developed from Stich's idea.



and believes that he will live only if he does not drink poisoned wine, then, *ceteris paribus*, X will not drink poisoned wine. Because such principles are close to being analytic truths definitive of rationality, they are far from being informative empirical generalizations.

CP's projective character, and its lack of the usual sort of empirical generalizations, suggests that it is unwarranted to call CP an empirical theory. However, it seems harmless enough to admit that CP is a theory of some kind. So long as we keep in mind the differences between CP and classical empirical theories, I have no objection to adopting the ubiquitous "theory" terminology. Given that we admit CP to be a theory, albeit of a special sort, our next question must be: are there any good reasons for thinking that CP is an inadequate theory?

Surely, all parties must acknowledge that CP works pretty well as an everyday system for explaining and predicting the behavior of normal, rational persons. We rely upon this system constantly, and it seldom fails us. One factor that philosophers have cited as an inadequacy of CP is the failure of belief attributions utilizing propositional that-clauses to index accurately the causal

roles of internal states. The most intuitive way to see the alleged problem is to note that the that-clauses utilized by users of CP to characterize the internal states of beliefs do not always capture unambiguously the way the believer conceives of his situation.

Take Kripke's example of the unfortunate Pierre, who thinks that 'London' and 'Londres' refer to two different cities (Kripke 1976). He believes that the city referred to by 'Londres' is pretty, but he believes that the city referred to by 'London' is not pretty. By using the familiar that-clauses of CP, we can attribute to Pierre, without evident mistake, both the belief that London is pretty and the belief that London is not pretty. Yet Pierre suffers no internal, psychological contradiction. The internal states that will actually explain his behavior and his reasoning are more finely individuated than that-clauses can accurately specify. CP thus seems to fail to capture the explanatorily-relevant psychological contents of beliefs with perfect accuracy.

But the fact that that-clauses fail to capture the psychologically relevant contents of beliefs with a perfect lack of ambiguity fails to show that CP is fatally inadequate. What is the purpose or function of CP,

anyway? It seems that the function of this theory of practice is the explanation and prediction of the normal behavior of ourselves and other persons who may properly be considered rational. And how often is it that the ambiguities latent in that-clause attribution cause any serious interference with this purpose? Not often at all. We can usually determine from the context what the psychologically relevant content of someone's belief is, even if we cannot assign a that-clause that perfectly pins down such content. The ambiguity of that-clause attribution is perhaps a minor inadequacy of CP.<sup>43</sup>

Churchland's definition of EM also implies that CP could be replaced by a neuroscientific theory. A critic might well inquire whether it is really possible for a neuroscientific theory, or any other kind of theory, to perform CP's function as well as, or better than, CP. Perhaps CP is disanalogous to other so-called "folk

---

<sup>43</sup> The inadequateness of CP has been pressed in another form. It has been argued that CP fails utterly to explain or predict the behavior of very young children, neurologically damaged persons, or persons with bizarre doxastic systems (Stich 1983: ch 4; P. Churchland 1986: 223). Indeed, in such cases we are hard-pressed to characterize the contents of mental states by means of that-clauses at all. But is this necessarily an indictment of CP? It is hard to see why. CP can fairly be taken to be a system for the prediction and explanation of the behavior of normal persons, old enough and similar enough to ourselves that we are comfortable treating them as rational. There is no reason to expect such a system to work in the case of abnormal, non-rational subjects.

theories" in that it explains facts that cannot be equally well, or better, explained by some other theory. The eliminativist argument that CP is replaceable, and likely to be replaced, seems to depend heavily upon the comparison between CP and other purported "folk theories." It is basically an argument by analogy: CP is similar to other folk theories that have proven false and been replaced; therefore, it is likely that CP too, will prove false and be replaced.

It seems to me, the claim that the entire notion of a folk theory is so vague that comparisons among various supposed folk theories are of dubious value. The many things that have been called folk theories are very different from each other. The argument for the elimination of CP based upon an analogy between CP and "other folk theories" that merit elimination strikes me, accordingly, as extremely weak. It seems that what CP says about propositional attitudes seems even more unlikely to prove false. Daniel Dennett (1987: 39) has argued that CP could not be replaced by any other theory because it captures certain unique and important generalizations. According to this line of argument, CP describes certain objectively real patterns or

regularities in the fabric of reality, that cannot be detected otherwise than by categorizing reality in intentional terms (by seeing persons as having states that refer to, or are about, their environment). Any explanatory framework other than CP misses something, according to this line of argument; CP is necessary in order to describe reality and in order to explain all the facts.

The argument that CP captures certain important generalizations and enables us to make otherwise impossible predictions is advanced in support of the prediction that no other theory will prove adequate to take CP's place. It does seem, then, that when we view creatures as rational, patterns and regularities in their behavior become visible that would not otherwise be detectable. Instead of merely seeing physical objects reacting to physical forces, we see episodes of inferring, perceiving, detecting, calculating, and other intelligent activities. Rational creatures, rather than just responding to stimuli, can respond to the meaning or significance of stimuli in the light of their own interests. Failing to take regard of this fact does, it seems, result in a significant loss of explanatory and

predictive power. No other sort of theory could replace CP. CP just is the conceptual framework in terms of which persons are rational beings or cognizers, and without this conceptual framework certain facts are inaccessible. Someone may think that it is just an empirical question whether CP turns out to be replaceable by some other theory or not. We must simply wait and see if future neuroscience, or some other future theory, turns out to be powerful enough to explain all that CP explains, and more. But I believe it is wrong to look at the issue in this way. The question is not an empirical one so much as a conceptual one. CP is the descriptive/explanatory framework that takes us to be rational persons and cognizers. Any significantly different theory could not explain the rational actions that CP describes and subsumes, because, without CP's concepts and vocabulary, there would be no rational actions to explain.

So far, I have argued that there are good reasons for thinking that CP is not an ordinary empirical theory; that it is not inadequate for its purposes, and that it could not be replaced by anything else. Churchland's definition of EM, as applied to CP, is dubious. When eliminativists actually argue for the thesis that

propositional attitudes deserve elimination, what they say generally has little to do with how well or how poorly CP works for its humble, everyday purposes. Arguments for EM tend to proceed from considerations having to do with the naturalistic reduction of theories and theoretical entities.

Could CP possibly prove false? Fodor has expressed very nicely the spirit behind the argument that CP could not be possibly proven false:

Even if [CP] were dispensable in *principle*, that would be no argument for dispensing with it ... What's relevant to whether commonsense psychology is worth defending is its dispensability *in fact*. And here the situation is absolutely clear. We have no idea of how to explain ourselves to ourselves except in a vocabulary which is *saturated* with belief/desire psychology. One is tempted to transcendental arguments: What Kant said to Hume about physical objects holds, *mutatis mutandis*, for the propositional attitudes; we can't give them up *because we don't know how to*. (1987: 9-10)

Indeed, there is something very odd and paradoxical about the idea that CP could prove to be false. What evidence could possibly show CP to be false? Recall that we are taking CP to be not only an explanatory and predictive calculus, but also the conceptual framework or descriptive vocabulary in terms of which persons are seen

as rational beings and cognizers. Whether CP could prove false is, accordingly, the question of whether it could turn out that persons are not rational beings, not cognizers, after all. It can seem that persons are just obviously rational beings, and that this is a truth too fundamental to be seriously questioned. Yet, we must acknowledge that to a certain kind of radical eliminativist it seems obvious that any theoretical framework, other than that of fundamental physics, could prove to be false. According to such an eliminativist, it could very well turn out that there were no such phenomena as rationality, intelligence, and cognition. The eliminativist claims that those terms derive their meanings from a theory that may be a thoroughly false description of reality. Perhaps, when we look at human beings, we ought to see physical particles responding to physical forces; perhaps that sort of description is the only true description. Perhaps, to look at human beings and to see episodes of perceiving, inferring, theorizing, and so on, is just wrong; the vocabulary in which these descriptions are couched may simply not be getting at any real phenomena.



I take it to be true that many different vocabularies, at many different theoretical levels, might all provide correct descriptions of reality; the radical eliminativist is one who takes it to be the case that only one vocabulary, that of fundamental physics, can give a true and correct description of reality. I defend the idea that CP couldn't possibly prove to be false, in the sense that we could not conceivably turn out not to be rational beings.

Quine has taught us that no theory taken in isolation is conclusively falsifiable, and that no theory is immune from revision (Quine 1951: 40-43). We can always save our favorite theory from elimination by altering some other part of the theoretical network. Any theory can, in principle, be revised or abandoned, or held inviolate. Let us suppose that Quine is correct about this. Then, if CP is a theory, what seems to make it different from other theories is that it is one we would be extremely reluctant to give up. Faced with giving up CP, or with giving up some other cherished theory, it seems we would give up the other theory.

## 4.5 CONCLUSION

By rejecting NCC we can in fact arrive at a sustainable, defensible and rewarding account of mental causation. The new conception of causation that has emerged is strengthened by a theory of intentional action that I will endorse in the last two chapters. A series of experiments (Knobe 2003a; Mele 2001; Malle & Knobe 1997; Mele & Moser 1994) demonstrate that our ordinary practice in attributing intentional action in particular cases, and our practice of attributing reason explanations, can actually be influenced by normative considerations. This result suggests that normative considerations may actually be playing a role in the concept of intentional action and reason explanation.

Our chief aim in Chapters Five and Six is, therefore, to present a convincing case for the conclusion that normative considerations actually play a role in the fundamental competence underlying people's causal attributions. Then, the widely held belief, one that mental causation should be understood as something like a scientific hypothesis, or the other that mental causation should be grounded on a purely naturalistic relation

between events, should be discarded. Our ordinary practices of attributing mental causation have an essential normative element - they are concerned not only with what is the case but also with what ought to be the case.

## CHAPTER 5

### INTENTIONAL ACTION AND NORMATIVE CONSIDERATIONS

While there is a disagreement among people concerning how to analyze the concept of intentional action, everyone seemingly agrees that the distinction between intentional and not-intentional<sup>44</sup> action plays an important role in our collective folk psychology.

According to the usual account we have some independent ground of what it means for an action to be intentional: an action is intentional when it is done for a reason. However I will show in this chapter that without taking moral considerations, the usual account

---

<sup>44</sup> In Mele and Moser 1994, they mention Harman's sniper (1976). In firing his gun, the sniper's position is knowingly informed to his enemy. Even though he does not intend to alert the enemy to his presence, he does seem to intentionally alert his enemy. In this case they say, he does

accidentally alert the enemy, it is natural to insist that he does not *unintentionally* alert the enemy. Such insistence does not entail, however, that the sniper *intentionally* alerts the enemy. There is a middle ground between A-ing intentionally and A-ing unintentionally. We locate 'side-effects actions' of the kind in question on that ground. In so far as such actions are not done unknowingly, inadvertently, or accidentally, they are not unintentional. In so far as the agent is not aiming at the performance of these actions, either as ends or as means to (or constituents of) ends, they are not intentional either. We shall say that they are *non-intentional*. (230-231)

Mele and Sverdlik (1996) also claim that there is a middle ground between unintentionally  $\Phi$ -ing and intentionally  $\Phi$ -ing, namely, non-intentionally  $\Phi$ -ing. I am not concerned with this issue in this work, though.

cannot cover all the categories of intentional actions. The thought that people are always starting with a judgment that an agent acted intentionally and then use it as input to a process that eventually yields, for example, a moral judgment, is ungrounded. The correct procedure needs in some cases to start with moral considerations and then use them to input a process that eventually yields a judgment that the behavior in question is intentional. This position is similar to the one that we saw in causal explanations.

According to the usual account causal explanations count as causal if they are grounded in causal relations. However, as argued earlier, the primacy of the explanatory practice over the ontological commitment reverses the usual account; explanations come first, such that an explanation is causal if we accept it as such. Here by reinterpreting the notion of causation we regain the causal efficacy of the mental. The problem raised by the Exclusion Argument, as we already saw, takes a wrong point of departure by always beginning with a metaphysical notion of causation instead of grounding the notion of causation on our explanatory practices. Likewise the usual account of intentional actions takes a

wrong point of departure when it always begins with the notion of intentional actions as actions done for reasons, therefore neglecting the point of grounding the notion of intentional actions on normative considerations.

In this chapter I will explain why we need sometimes to reverse the usual account of intentional actions in order to cover all the categories of intentional actions. I will first argue for this point by examining some cases on intentional actions, which show that the moral qualities of the outcome of a behavior strongly influence people's judgments as to whether that behavior should be considered intentional. Here the most important point to notice is that people not only rely on their judgments of action's being intentional to make moral judgments, but the contrary is true as well - i.e. sometimes people's moral judgments influence their ascriptions of intentional action.

In order to show this point I will examine some of the views that have been forwarded in the philosophy of action literature concerning intentional actions. That means, I set the stage by examining some of the problems associated with the concepts of intentional action that are frequently discussed in the literature on the

philosophy of action: lucky actions and unintended side effects.

In section 5.1 I will discuss the so-called Standard Account of intentional action and its difficulties. This discussion is closely related to the discussion of section 5.5, where what I call the Simple View is introduced. I will pay particular attention to the view concerning the relationship between skill, control, foresight and intentional actions (section 5.2), and between unintentional side effects and intentional actions (section 5.3) with regard to the Standard Account. I will then provide an explanation of understanding intentional action by invoking and distinguishing motivating reasons from normative reasons. Finally, I will show that there is a gap between what is required for intending to  $\Phi$  and what is sufficient for intentionally  $\Phi$ -ing by rejecting what I shall call the Simple View. I elicit, by rejecting the Simple View, a *theoretical ground* for taking normative perspectives in dealing with the concept of intentional actions.

## **5.1 THE STANDARD ACCOUNT OF INTENTIONAL ACTIONS AND ITS DIFFICULTIES**

The common starting point for theories of intentional action is the observation that intentional action is action done for a reason. In her groundbreaking work *Intention* (1957), Elizabeth Anscombe expresses the thought as follows:

What distinguished actions which are intentional from those which are not? The answer that I shall suggest is that they are the actions to which a certain sense of the question 'Why?' is given application; the sense is of course that in which the answer, if positive, gives a reason for acting." (Anscombe 1957: 9)

I will characterize this account as the "Standard Account"<sup>45</sup>:

[The Standard Account]: An agent  $\Phi$ -es intentionally if and only if she  $\Phi$ -es for a reason.

The 'for a reason' locution implies that what the agent did can be explained by citing her reason for acting. The explanation, according to this account, will be an explanation of a certain sort; it will be an explanation of what the agent did from her point of view.<sup>46</sup> Thus the

---

<sup>45</sup> Audi also claims that all actions done for a reason are intentional (1986: 514).

<sup>46</sup> Anscombe held, following Wittgenstein, that to give a reason for an action is not to provide a causal explanation of it. Anscombe relied on the justifying function of reasons, as did philosophers



Standard Account starts out with the assumption that we have some independent notion of what it means for a behavior to be performed for a reason and that we can use this notion to arrive at an understanding of the distinction between intentional and not-intentional behavior.

We are entitled to infer that Davidson also subscribes to the Standard Account. According to him someone is the agent of all events for which there is at least one true description under which he did something intentionally (1971: 46). In another essay Davidson indicates that acting intentionally implies acting for a reason. He puts it this way:

[Suppose that the agent's] action is intentional. We must therefore be able to abstract from his behavior and state of mind a piece of practical reasoning the conclusion of which is, or would be if the conclusion were drawn from the premises, that the action ... performed is desirable.<sup>47</sup> (1969: 32-33)

In other words, in order for an action to be intentional, the agent must have in mind a reason, or reasons, which rationalize her action as to she performs it. I do not

---

such as Melden. Giving a reason helps us understand why the agent did what she did. I will take it to be true, however, following Davidson, that the reason for an action is its cause.

<sup>47</sup> The desirability here should be from the agent's point of view.

think that Davidson means that the agent must consciously work through a piece of practical reasoning; instead, what is necessary is that the reason be present in her mind, present from her point of view, and that it should play a role in how and why she acts.<sup>48</sup>

However the concept of intentional action comes sometimes apart with reason-explanations. The Standard Account is challenged by some cases of extraordinary luck. The case I will examine in section 5.2 is the case where in order for an agent to intentionally  $\Phi$ , her  $\Phi$ -ing must be the result of a certain amount of skill or control. In other words, the claim is that an agent cannot intentionally  $\Phi$  if her  $\Phi$ -ing was primarily the result of luck. In cases where the agent seems not to have enough control over the effect of the behavior, people do not use the same criteria to decide whether the effect of the behavior was intentional. Therefore some people claim that an agent cannot intentionally  $\Phi$  if her  $\Phi$ -ing was primarily the result of luck. The problem is that the

---

<sup>48</sup> Davidson's position is in fact weaker than the Standard Account since Davidson seems to be silent about the issue as to whether everything done for a reason is intentional.

Standard Account of intentional actions is not able to deal with this case.

Causal deviance is another similar challenge that is traditionally raised against the Standard Account. There are cases such that an action was done for a reason, however it seems not be taken as intentional because of causal deviance.<sup>49</sup> In Chapter Six I will show, as in the cases involving skill/luck, that the moral qualities of the outcome of a behavior in the cases of causal deviance strongly influence people's judgments as to whether that behavior should be considered intentional.

An unintended but foreseen side effect also gives a counterexample to the Standard Account. The unintended side effects are not among the things agents can be said to bring about intentionally because the effects were not done for a reason. I will argue for the claim that the account of intentional actions, in some cases, will be affected by moral considerations. Now let's take a look at those challenges in turn.

---

<sup>49</sup> I will deal with the problem of causal deviance in Chapter Six because the problem is closely related to mental causation debate itself.

## 5.2 THE CHALLENGE TO THE STANDARD ACCOUNT: SKILL/LUCK

The Standard Account is challenged by some cases of extraordinary luck. While any  $\Phi$ -ing that involves too much luck to be regarded as intentional, it is possible for  $\Phi$ -ing to be explained using reasons. What this means is that there are actions, done for reasons, that are not intentional.

Consider a case in which an agent is trying to perform a behavior and actually does succeed in performing that behavior. And now suppose that the agent didn't really have the skill to perform that behavior in any reliable fashion, so that ultimately the agent only manages to succeed through sheer luck. Harman gives an example involving a sniper who shoots a bull's-eye (Harman 1976: 433-34). The sniper is trying to shoot and actually does shoot the bull's-eye, but only succeeds in performing the behavior through sheer luck. The point in this case is that the sniper didn't really have control over the result; success in shooting the bull's eye is not the result of any relevant skill or control on the sniper's part. The sniper's success is through luck. In

this case, our intuition strongly says that his shooting is not intentional. The question is simply whether people use the same rule to determine whether a behavior was performed intentionally as they use to determine whether a behavior was performed for a reason, since according to the Standard Account people determine whether a behavior is intentional by examining whether it is performed for a reason. And the answer is, if the above intuition is right, they don't; an agent cannot intentionally  $\Phi$  if her  $\Phi$ -ing was primarily the result of luck, a counterexample for the Standard Account.

What this shows is that it seems intuitively plausible that if an agent has no control over the result of her  $\Phi$ -ing, or she luckily manages to  $\Phi$ , we should not say that she intentionally  $\Phi$ -es. From this consideration, some philosophers, for example, Mele and Moser (1994), say that when luck plays a role in the success of an attempt at  $\Phi$ -ing, the  $\Phi$ -ing is generally deemed too coincidental to count as intentional, and conclude that a relevant amount of skill or control is a necessary condition for an action to be performed intentionally: an intentional action cannot be the result of luck.

However the issue is complicated, since even though  $\Phi$ -ing was not the result of any relevant skill on the part of the agent, there are related cases where people often judge that an agent  $\Phi$ -ed intentionally. In order to show this Harman gives another example involving a sniper who shoots a soldier. In this case, however, the situation changes when the sniper succeeds in shooting the soldier even though it is performed by luck. People's intuition is saying that the shooting, if it succeeds, is intentional. Harman claims:

The reason why we say that the sniper intentionally kills the soldier but do not say that he intentionally shoots a bull's-eye is that we think that there is something wrong with killing and nothing wrong with shooting a bull's-eye. (Harman 1976: 433-34)

What the above case shows is that in some cases our concept of intentional action is not sensitive to considerations of skill, luck, and control. This case alone shows Mele and Moser wrong; we should reject any analyses of the ordinary concept of intentional action that has skill, control or the absence of luck as a necessary condition. This case also shows that the concept of skill, luck, control does not help to analyze

the concept of intentional action. Instead, in some special cases, we seem to need to look at the moral status of the result of agent's  $\Phi$ -ing itself. The two examples are structurally similar. However while in the former case, we are not able to attribute, for example, blame to the agent in question, in the latter we want to ascribe blame: in the former case our intuitions tell us that luckily bringing-about is not sufficient to justify the attribution of intentionally bringing-about; in the latter our intuitions say that luckily bringing-about is sufficient for intentionally bringing-about. The average person's intuition about the cases concerning the features of skill, luck, and control seems to sometimes depend on the moral status of the behavior itself. Namely moral considerations play a role in people's intuitions whether an agent's behavior is intentional. In this way, normative considerations come in the talk of intentional actions, which is the subject of Chapter Six.

### **5.3 THE CHALLENGE TO THE STANDARD ACCOUNT: UNINTENDED SIDE EFFECTS**

There has been a great deal of controversy in the philosophical literature about the role that trying and foresight play in the concept of intentional actions. Some philosophers think that trying is a necessary condition for intentional action (Adams 1986; McCann 1986); others argue that a certain kind of foresight can actually be sufficient even in the absence of trying (Ginet 1990). The distinction between these two views comes out most clearly in cases of what might be called unintended but foreseen side effects. An outcome can be considered an unintended foreseen side effect when (1) the agent was not specifically trying to bring it about but (2) the agent chose to do something that she foresaw would involve bringing it about. If trying is a necessary condition for an action being intentional, the agent did not bring about the side effect intentionally. By contrast if foresight is sufficient for an action being intentional, the agent brought about the effect intentionally. In the latter case then an unintended foreseen side effect gives a counterexample for the Standard Account; the unintended side effects are not among the things agents can be said to bring about



intentionally because the effects were not done for a reason.

Let me consider the following Strategic Bomber case of an unintended but foreseen side effect:

[The Strategic Bomber (SB)]: SB intends to bomb a munitions plant as a means to his ultimate end of winning the just war, knowing that there is a school next door, therefore foreseeing that his bombing will bring about civilian deaths as an unwanted but unavoidable side effect. (Bratman 1987: ch. 10)

SB acts in pursuit of a certain end - he wants to win the just war - and on the basis of a certain belief - that he can win the war by bombing a munitions factory. What he does can be explained in the "for a reason" sense under descriptions like "bombing a munitions factory." We can therefore say he blows up the factory intentionally. His behavior, however, cannot be rationalized under the description, "killing the civilians," since killing the civilians cannot be explained as something done for a reason. If the Standard Account is right, we cannot say SB killed the civilians intentionally, because there is no explanation of the 'for a reason' variety of his killing them.

Our intuition, however, says that SB seems to be, for example, responsible for killing the civilians. Our

intuition strongly suggests that he killed them intentionally. On the Standard Account, however, this is not a conclusion we are entitled to reach. The demand of dealing with the cases of unintended side effects conflicts with the Standard Account of intentional action. On the Standard Account the category of intentional actions is quite narrow. If our intuition is right, then any criterion for identifying whether an action is intentional or not would have to deal with the above case. What this seems to suggest is that whether I  $\Phi$ -ed something intentionally depends, sometimes, on whether the thing I  $\Phi$ -ed had good or bad effects, though I did not intend to bring them about. The subject of section 6.2 is to show that the account of intentional action, in this unintended side effect case, will be affected by moral considerations.

#### **5.4 NORMATIVE/MOTIVATING REASONS**

Let's take a look at the two notions of reasons, normative and motivating reasons.<sup>50</sup> This is a distinction

---

<sup>50</sup> One might wonder which of these is at issue in Davidson's account of reasons for actions. Davidson seems to want to use the technical

between reasons that merely justify a certain type of action, and reasons that explain why an agent performed such an action. In the former case, we might speak of "a reason for a certain sort of action," and in the latter case "a reason why an agent performed such an action."

The notion of normative reason is one that we consider when we speak in favor of, or against, a course of action. When we deliberate about what to do, we reflect on such considerations as they bear on possible action, and if they show that an action should be done, we are bound, if we are rational, to act on them. That such consideration can obligate us to act is why we call them reasons. Sometimes by the expression "an agent's reasons" we are concerned with the normative claims of a theory of rational action, so that we might say, for example, that all agents have good reasons for

---

notion of a primary reason to speak of the motivating sort of reasons, those that are explanatory. However, Davidson not only speaks of "a primary (motivating) reason why an agent performed an action," but also of "a primary reason for an action." In stating his first necessary condition concerning primary reasons, Davidson speaks of "a primary reason why an agent performed an action," which clearly indicates that what is being characterized is the sort of reason that explains why an agent performed such an action. Davidson's second necessary condition, a primary reason for an action is its cause, is certainly intended as a correlative condition to the first one and concerns these explanatory reasons as well, but the terminology he uses in stating the second condition fails to make this clear. Maybe he has this in mind when he says the second necessary condition: "R is a primary reason why an agent performed the action A only if R caused A."

cultivating their talents. Maybe what this means is that cultivating one's talent serves as a means to the various ends that agents might pursue. It does not follow from this, of course, that all agents want to cultivate their talent or even that they would agree that cultivating their talent is a good thing. When concerned about the relation between reasons and actions, to speak of an agent's reasons is to speak of reasons the agent actually holds, whether these reasons conform to our normative theory of rational action or not. The reasoning in question need not meet the standards of our normative theory of rational action: the standards which specify which ends agents ought to pursue and which actions are the most reliable or reasonable means to those ends.

Our normative reasons do not only obligate us, but motivate us if we are rational, and this talk of motivation brings us to the notion of a motivating reason. One way to understand motivating reasons is to link them with the specific question type that they typically answer, "Why did an agent  $\Phi$ ?". Consequently, it is sometimes said that a motivating reason is a reason why.

Thus, philosophers often claim that there are two different sorts of reasons: reasons for action that have a normative bearing on things we might do and reasons that explain why we do those things. Yet if we acknowledge that agents sometimes act for reasons, i.e. act on the basis of normative considerations, then it seems that they are motivated by those reasons. Indeed, to say that rational agents must have the capacity to act for reasons is to say exactly that normative reasons must be capable of motivating them, i.e. of being motivating reasons.

### **5.5 THEORETICAL GROUND FOR NORMATIVE CONSIDERATIONS**

In this section I show that the so-called Simple View is false. I will characterize the Simple View as follows:

[The Simple View]: One intentionally  $\Phi$ -ed only if one intended to  $\Phi$ .

The Standard Account of intentional action entails the Simple View. The importance of discussing the Simple View on our purpose is that by showing the falsity of the

Simple View we have a *theoretical ground* for normative considerations in dealing with the concept of intentional actions.

Philosophers have tried to give an account of the relationship between "intentionally  $\Phi$ -ing" and "intend to  $\Phi$ ." According to the Simple View in order for an agent to  $\Phi$  intentionally, she must have intended to  $\Phi$ ; one is entitled to infer, from the fact that an agent intentionally  $\Phi$ -ed, that she intended to  $\Phi$ .<sup>51</sup> On this view there is no difference in scope between the intended and the intentional.

The Standard Account of intentional action entails the Simple View. For if one accepts the Standard Account of what is done intentionally, there will be no room left over for a distinction between the intended and the intentional action.<sup>52</sup> On the Standard Account it makes no sense to speak of doing something intentionally when what the agent does is contrary to what he desires.

Audi puts forth the example of the poor shooter who attempts to hit a bull's eye on a distant target (Audi,

---

<sup>51</sup> Adams (1997; 1986) and McCann (1986) hold this view.

<sup>52</sup> The simple view, however, does not necessarily entail the Standard Account, since it is possible to associate the intended with the intentional and to associate neither with what is done for a reason.

1973: 401). Much to his surprise, the shooter hits the target, the bull's eye. Davidson offers a similar case in which a person tries to make ten carbon copies on a typewriter while doubting that it can be done (1978: 92). Again, much to our typist's surprise, each of the copies is successfully made. It is strongly intuitive to some, including Audi and Davidson, that in both of these examples the agents intentionally  $\Phi$ -ed. If a strong belief requirement, the requirement that S intends that  $p$  only if S believes that  $p$ , is placed on intending such that intending to  $\Phi$  implies believing that one will  $\Phi$  and if there are cases where one intentionally  $\Phi$ -es even though she doubted that she was  $\Phi$ -ing at the time, then the Simple View must be false.

Bratman (1987: 113-116) gives a more direct argument against the Simple View. In the words of Bratman, "The Simple View supposes that there must be a tight fit between what is done intentionally and what is intended" (119). His argument involves an example of a video game in which the player is able to play a missile target game with each hand. The game is constructed in such a way that one wins if one hits one of the two targets. One

cannot, however, hit both targets or else the game will shut down. When one hits one of the targets, it is clear, according to Bratman, that one has done so intentionally. Thus, if the Simple View is correct, one must have intended to hit the target. The problem, says Bratman, is that one must have intended to hit the other target as well. However one cannot have so intended because one's intentions would not be consistent - they would involve one in a criticizable form of irrationality. Yet according to Bratman, "it seems clear that I need be guilty of no such irrationality: the strategy of giving each game a try seems perfectly reasonable" (114). Thus, the Simple View, says Bratman, must be false.

There would be gap between what is required for intending to  $\Phi$  and what is sufficient for intentionally  $\Phi$ -ing if the Simple View is false. And I think the arguments against the Simple View are persuasive. Then the intentional and the intended must be pulled apart. What this means is that the boundaries of intentional actions are sometimes derived from things that agents do not intend to do. Now because of the gap, we must be, in some cases, able to treat the case of  $\Phi$ -ing intentionally



as a non-psychological notion. The concept of an intention to  $\Phi$  is entirely a "psychological" concept. Intentions are connected with motivating reasons. You intend to  $\Phi$  something only if you view yourself as having a reason to  $\Phi$  it. Intentions are a species of reason for acting in the explanatory sense. Since normative judgments, seen from third-party perspective, can apply irrespective of the psychological state of the agent, we will find intentional action applicable in many cases in which the agent does not do what she does 'for a reason' in the explanatory sense of that phrase.<sup>53</sup>

What this consideration shows is that the criteria for intentional action must be wide enough to include the normative perspectives of third-person point of view as well as the psychological perspectives. I take this as providing a theoretical ground that we should take normative considerations of third-person point of view in dealing with the concept of intentional actions. On the one hand, an agent does something intentionally if doing it was her reason for doing what she did, namely the

---

<sup>53</sup> Third-party perspective because it does not matter whether or not the reasons in normative judgments provide the agent with a motive to perform the action.

consideration that moved her to perform the action. Of course, this is an explanation consistent with the Standard Account of intentional action. On the other hand, people in some cases judge that an agent does something intentionally by taking normative considerations on the basis of third-party concerns, rather than on the basis of how things looked from the perspective of the agent.

What is done intentionally should, in specific cases, accommodate the demand that normative considerations make of action, while intending to  $\Phi$  captures the psychological perspective we adopt when we are concerned to explain what an agent does in terms of her reasons for acting. The notion of intention is captured by agent's explanatory reasons, but the intentional is, in some specific cases, turned toward the normative therefore is not wholly understood by considering only explanatory reasons. While what is intended sides with explanatory reason, what is done intentionally sides partly with normative reason.

## 5.6 CONCLUSION

Everyone seemingly agrees that the distinction between intentional and not-intentional action plays an important role in commonsense psychology (CP). People have classified behaviors as intentional or not-intentional by trying to give an explanation of the conception of intentional actions without considering more important questions such as normativity. Some philosophers, who hold the view that CP is best understood as a tool for predicting and explaining behavior, suggest that CP is a kind of proto-science. And they appear to feel that normative considerations just couldn't be playing a fundamental role. The view that CP is a kind of proto-science is, as I argued in Chapter Four, ungrounded. I take it that CP is a practice. I argued that our explanatory practice should guide our ontological commitments.

The solution for the problem generated by the Standard Account in explaining intentional action is to rethink the notion of intentional action. In this chapter I show some hints that moral considerations have an impact on people's judgments of intentional action. We will see that people's concept of intentional action is bound up in a fundamental way with evaluative questions.

I will show, by defining the concept of intentional action, that folk ascriptions of intentional action are sensitive to normative considerations, not limited to moral considerations. Based on this claim, I will argue that normative considerations play some role in solving the problem of mental causation debate.

## CHAPTER 6 NORMATIVE CONSIDERATIONS AND MENTAL CAUSATION

The task of defining intentional action has given rise to heated debates in contemporary philosophy. As the previous chapter hinted, however, it is not enough to fully understand the phenomena of intentional action by explaining and analyzing only the agent's reasons that accompany each type of action. We saw some hints that normative considerations have an impact on people's judgments of intentional action.

In Chapter Five, I argued for this point by invoking a theoretical ground for us to include normative considerations of third-person point of view in dealing with the concept of intentional actions. Then, the definition of intentional action should be bound up with evaluative questions because the concept of intentional action should be sensitive to normative considerations. The criteria for intentional action must be wide enough to include the normative perspectives of a third-person point of view as well as the psychological perspectives.

In this chapter I will sharpen this idea by looking at recent empirical research and propose to understand

intentional action in terms of both motivating and normative reasons. I will provide a novel conception of intentional action by distinguishing normative reasons from motivating reasons. The definition should be reflected on both reasons: on the one hand, an agent does something intentionally if they were her reasons for doing what she did, namely the consideration that moved her to perform the action, consideration consistent with the Standard Account of intentional action; on the other hand, we say normative considerations play a role in people's intuitions whether an agent's behavior is intentional.

The proposal recommends itself as being capable of dealing with many problems, including the problems raised by unintended side effects and lucky actions. More importantly, the proposal is able to deal with the problem of casual deviance and consequently is promising in that it avoids epiphenomenalism of mental properties. While the solution for the problem generated by the Standard Account in explaining intentional action is to rethink the notion of intentional action, the causal efficacy of the mental is to be guaranteed by reinterpreting the notion of causation.

In section 6.1 I provide my definition of intentional action. The criterion for intentional action I am suggesting straddles the psychological and the normative perspectives in order to deal with problematic cases. In section 6.2 I address the problem of unintended side effects. An empirical research performed by Knobe (Knobe 2003a) shows that people's intuitions are influenced by the moral qualities of the side effect itself. This intuition is reflected in my definition. Section 6.3 deals with the cases involving luck. In this case normative considerations also play a role. I show one merit of my definition; it explains people's different intuitions on whether an agent performs a behavior intentionally when the result seems to be due to luck. I also show that my definition confirms the result of Chapter Five that skill and control are not necessary components of the concept of intentional action. Section 6.4 is also dedicated to showing that the moral qualities of the outcome of a behavior in the cases of causal deviance influence people's judgments as to whether that behavior should be considered intentional. We will also see the merit of my definition in being able to deal with people's different intuitions on whether an agent

performs a behavior intentionally when the result seems to be due to causal deviance. In section 6.5 I will examine the difference between mental explanation and naturalistic explanation. From this consideration I claim that because of the justificatory factor in dealing with intentional action, it is difficult to see how NCC can be true. I argue that NCC is an error due to confusing a mental explanation with a purely naturalistic explanation between events. I further claim that this insight works nicely in the case of causal deviance. Finally I argue that NCC is just the result from supposing that there is no gap between explanatory reason and justificatory reason.

## **6.1 INTENTIONAL ACTION**

The distinction between intentional and non-intentional actions plays an important role in commonsense psychology (CP). For example, in ordinary situations, the question of whether or not an action was performed intentionally can make a big difference in how we respond to it. However there is disagreement among philosophers as to how to analyze and define the concept



of intentional action. The issue of the debate is whether moral/normative considerations do affect our application of the concept of intentional action. Some people claim that moral considerations should not act on our ascriptions of intentional action (Butler 1978; Mele and Sverdlik 1996). On this view, while we may correctly appeal to the fact that an action is intentional in order to determine whether the agent in question is morally responsible, the converse is not the case; attributions of responsibility should not influence our ascriptions of intentional action. Others (Bratman 1987; Harman 1976; Knobe 2003; 2004; Nadelhoff 2004) claim that the ascriptions of intentional action are intimately bound up with moral considerations. It may, at first, seem strange to take an account of moral considerations as a relevant factor as to whether the agent performed the action intentionally. However, the latter view has now received support in the philosophical literature.

I gave, in the previous chapter, some hints that with regard to the relationship between unintended side effects, skill/luck and intentional action people's intuitions are influenced by the moral status of the behavior. I also provided a theoretical ground to include

normative perspectives in dealing with the concept of intentional actions. This will be confirmed by people's intuitions on the concepts of intentional action, which is the subject of the next section.

Here I will provide a novel conception of intentional action. The conception that considers both explanatory and normative perspectives in dealing with the concept of intentional action, I argue, recommends itself as being capable of solving problems generated by the Standard Account of intentional action with regard to unintended side effects and lucky actions. More importantly for our purposes, however, it provides a way of looking at the mental causation debate by successfully dealing with causal deviance problems. The conception is as follows:

[Intentional action] An agent's  $\Phi$ -ing is intentional iff either (i) it is done for her motivating reason (if it is not the case of luck or causal deviance) or (ii) the fact that certain consequences would occur was a justifying reason not to perform the action.

The definition pays close attention to the normative considerations as well as motivating reasons. On the one hand, an agent's  $\Phi$ -ing is intentional if it was done for

her motivating reason, namely, the consideration that moved her to perform the action. On the other hand, an agent's  $\Phi$ -ing is intentional if from a third-party perspective, the fact that the consequence would occur was a "reason" not to perform the action, whether or not the reason in the latter sense was one that weighed with the agent as supplying a motive not to perform the action.<sup>54</sup> The notion of "reason" in this account thus alternates between an "explanatory" and a "justificatory" sense.

The difficulty in trying to provide an account for intentional actions stems from the task of harmonizing the two different perspectives, the psychological and the normative points of view. However, the definition I provide successfully deals with the difficulty. The former perspective comes in when we are concerned with understanding what led to a person to do something. The condition (i) reflects this perspective. In this case we focus on how things looked from *the agent's point of view*, and in particular, we look for an explanation in terms of what the agent thought she was accomplishing in so doing.

---

<sup>54</sup> The agent need not have been aware of the considerations.

When we are concerned with whether an action is intentionally done in some specific cases, however, we need to consider a broader standard than we did when we adopted the explanatory standpoint. Here the normative aspect comes in. The problem is that the broader standard, namely, the *justificatory standpoint* we adopt when we focus on this wider class of doings cannot be imposed on the basis of the explanatorily motivational standpoint the agent could have of what she did. The justificatory standpoint cannot rest on features which are psychological or motivational to the action, but rather must be imposed from outside. The condition (ii) reflects just this perspective. The "from the outside" perspective may happen to match with the perspective that weighed with the agent as supplying a motive not to perform the action but we have no reason to expect that the "from the outside" perspective is on the same ground as the explanatorily motivational perspective. Saying the consideration is a reason against performing the action is a claim of quite a different sort from saying it is a reason I regarded as weighing against my action. The third party consideration has a very different status from the agent's "internal" considerations.

To say that there is a justifying reason to  $\Phi$ -ing is to say that:

[T]here is some normative requirement that she  $\Phi$ 's, and ... that her  $\Phi$ -ing is justified from the perspective of the normative system that generates that requirement. (Smith: 95)

The perspective of generating those requirements may be diverse: it would be from rationality, prudence, or morality. Here I am not concerned the issue of whether moral perspective can be reduced to rationality perspective. All I claim here is that the perspectives depend on which societies we live. Therefore there is a justifying reason not to buy a lottery ticket if buying a lottery ticket is banned in the society, and there may be no justifying reason, for example, in an amoral society, not to kill an innocent person. This is the reason that my criterion for intentional action is not limited to just moral considerations but expanded to normative considerations.

## 6.2 UNINTENDED SIDE EFFECTS REVISITED

We saw in Chapter Five that people's intuitions about the cases concerning unintended side effect sometimes seem to depend on the moral status of the side effect itself. Namely moral considerations play a role in people's intuitions whether an agent's behavior is intentional. In this section I provide a result from a recent research to support this point, and take the result of the research as an empirical ground for us to include normative considerations with regard to intentional actions.

According to the result of Knobe's research (Knobe 2003a) people's intuitions appear to be influenced by the moral qualities of the side effect itself. According to this research people seem to be considerably more willing to say that the agent brought about the side effect intentionally when they regard that side effect as bad than they are when they regard the side effect as good.

Knobe (2003a) presents data that are taken to support this view. Knobe's data show an asymmetry in people's judgments. In a case of the side effect when people are asked whether the agent brought about the outcome intentionally, they are more inclined to judge that the agent did bring about the outcome intentionally,

if the outcome was perceived as causing a harm. There is an asymmetry because people are not inclined to see an agent's action as intentional if the outcome is perceived as causing a benefit. This idea is best understood by looking at the following examples that Knobe gives:

[Example 1] The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment.' The chairman of the board answered, 'I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was harmed. (2003a: 191)

[Example 2] The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, and it will also help the environment.' The chairman of the board answered, 'I don't care at all about helping the environment. I just want to make as much profit as I can. Let's start the new program.' They started the new program. Sure enough, the environment was helped. (2003a: 191)<sup>55</sup>

Now Knobe invites us to ask whether the chairman of the board intentionally harms the environment in the first example, and intentionally helps the environment in the second example. By using the above examples, Knobe wants

---

<sup>55</sup> Methodological objections may be raised against Knobe's results. I will not pursue them here. It is sufficient for my purposes that the results themselves, were they pursued in thought-experimental fashion, suggest robust intuitions.

to show us that the moral qualities of the outcome of a behavior strongly influence people's judgments as to whether that behavior should be considered intentional, and actually he concludes that the result indicates that people's concept of intentional action is influenced by moral considerations. The experiment shows that people are more likely to judge that a morally negative action or side effect was brought about intentionally than they are to judge that a structurally similar action or side effect that is morally positive was brought about intentionally.<sup>56</sup>

We cannot claim credit for good things we do that we merely foresee will follow from our actions; in the second case the chairman of the board does not seem to be able to claim the beneficial effect. The natural thought, then, is that the chairman did not bring about the effect intentionally. However, we must be held responsible for the bad effects of the actions we foresee. In the first example the chairman can be blamed for the effect that he

---

<sup>56</sup> People's judgments on whether non-side effect actions are intentionally done are sensitive to positive moral considerations in a way that their judgments of side effect actions are not. In the case of unintended side effects we would need to explain why negative but not positive moral considerations affect people's judgments concerning action's being intentional. This is also one of the reasons that I gave the definition of intentional action either (i) or (ii).



foresees but not intended; he brought about the effect intentionally. What this means is that whether I did something intentionally depends, sometimes, on whether the thing I did had good or bad effects, though I did not intend to bring them about. This shows that the account of intentional actions, in special cases, will be affected by moral considerations.<sup>57</sup>

People's intuition regarding the example is reflected in the definition of intentional action I gave in the previous section. If the effect is the case of unintended, but foreseen side effect, we do not look at the agent's motivating reason to decide whether the effect of  $\Phi$ -ing in question is intentional. Instead we need to look at the fact that certain consequences would occur was a justifying reason not to perform the action.

### 6.3 SKILL/LUCK REVISITED

People's intuitions about the cases involving luck are similar to the cases involving unintended side effects concerning the issue of an action's being

---

<sup>57</sup> Then this is a counterexample to the Simple View, a view that in order for an agent to  $\Phi$  intentionally, she must have intended to  $\Phi$ .

intentional. Like the cases we have seen in dealing with the problem of unintended side effects, normative considerations also play a role in the cases of lucky actions in determining whether an agent's behavior is intentional.

We saw in Chapter Five that there was a problem in the Standard Account of intentional action in explaining the actions done with regard to skill/luck. People's intuition regarding Harman's sniper examples is reflected in the definition of intentional action I gave in the previous section. If an agent's  $\Phi$ -ing is a case of luck, we do not look at the agent's motivating reason to decide whether her  $\Phi$ -ing in question is intentional. Instead we need to look at a justifying reason not to perform to  $\Phi$ .

Consider the case of winning a lottery ticket. Even though an agent really desires to win the lottery and she tries to win and actually does win the lottery, people would not say "she won the lottery intentionally," because the success of winning the lottery is through sheer luck. Winning the lottery is not the result of any relevant skill or control on the agent's part. The agent didn't really have control over the result of the lottery.

People's intuitions tell us that if an agent has no control over the result of her  $\Phi$ -ing, or she luckily manages to  $\Phi$ , we should not say that she intentionally  $\Phi$ -es.

Now the definition of intentional action I gave does not have any trouble in dealing with this intuition. Just ask whether there is any justifying reason not to win a lottery ticket. If the answer is "yes," the agent won the lottery intentionally, If "no," then the agent did not win intentionally. And I can claim with confidence that there seems to be no justifying reason not to win a lottery ticket.

Let us examine the point in more detail by taking a look at a problem that has been provoked a great deal of controversy. It is the Analysis Problem No. 16, raised by Ronald Butler. The problem is the following:

If Brown in an ordinary game of dice hopes to throw a six and does so, we do not say that he threw the six intentionally. On the other hand if Brown puts one cartridge into a six-chambered revolver, spins the chamber as he aims it at Smith and pulls the trigger hoping to kill Smith, we would say if he succeeded that he had killed Smith intentionally. How can this be so, since in both cases the probability of the desired result is the same?  
(Butler 1978: 113)

In order to solve the Analysis problem, we need to show why we refer to the former as an instance of non-intentional action, and the latter as an instance of intentional action. What explains the difference of people's intuition for these two structurally identical cases resulted from the different moral status of the two cases. The intuition says, as we saw in Chapter Five, people are more likely to judge that a morally negative action or side effect was brought about intentionally than they are to judge that a structurally similar non-moral action or side effect was brought about intentionally. The difference between Brown's rolling a six and his shooting Smith is that while nothing is wrong in the former, something is wrong in the latter. This difference explains the intuition that Brown did not intentionally roll a six whereas he did intentionally shoot Smith, even though his chances of success and his relevant control over the outcome are the same in both cases.

The definition I gave explains this intuition in Brown's shooting case. In the event that the agent, from the third-party perspective, has a reason not to bring

about Smith's death and yet she brings about his death, then, even though the killing was due to luck, we should judge that the agent brought about Smith's death intentionally. Even though the agent's rolling a six in a dice game is the same in chance of success as the case of shooting Smith, people do not say that the agent brought the effect out intentionally. Of course in this case the effect is not the result of any relevant skill on the part of the agent, and there is no problem of dealing with this case since it does not in any way conflict with normal people's intuition. In order to use the definition of intentional action I gave, however, we need to ask the following questions, "is there any justifying reason not to roll a six?" and there seems to be no justifying reason not to roll a six in the dice game. Then the action in question is not intentional.

One merit of my definition is the fact that it explains the different intuitions on whether an agent performs a behavior intentionally when the result seems to be due to luck. I use the word "seems" because people's intuitions vary on whether the case in question as one involving luck or not. According to Peacocke (1985), an agent who makes a successful attempt to hit a

croquet ball through a distant hoop *intentionally* hits the ball through the hoop even though the chances of hitting are extremely low. Some people, including me, do not agree with Peacocke. The possibility of the disagreement shows that sometimes it is not clear whether the case should be dealt with the lucky action.

Now the definition of intentional action I gave does not have any trouble dealing with this intuition. For example, consider Davidson's typewriter example that we saw in Chapter Five. If someone says that the typist's action is not intentional, as opposed to most people's intuition, my definition is able to follow her rationale; she is dealing with the result of the agent's action as being involved with luck. According to my definition if it is the case with luck, we need to ask whether there is a reason not to make ten copies, and the answer seems to be "no," therefore the action is not intentional.<sup>58</sup>

However, people's intuition strongly suggests that the agent intentionally made the ten carbon copies. What this means is that the case in question is not a case with

---

<sup>58</sup> The case, in fact, however, need to be analyzed in the following way: whether there is a reason not to make ten copies, and the answer, here, is "yes" because the agent actually doubts that she will do.

luck; in this case people put more interest in the question "whether it is done for the agent's motivating reason" than the question "whether there is a justifying reason not to perform the action." Peacocke's intuition that the agent did hit a croquet ball intentionally seems to result from his emphasis on the fact that it is done for her reason, and so it is intentional, than on the fact that the case in question is one where luck is involved.

The solution of the Analysis Problem and Harman's sniper example, we saw in Chapter Five shows, that skill and control are not necessary components of the concept of intentional action. I showed that my definition of intentional action is able to deal with these cases, where normative considerations sometimes trump considerations of skill, luck, and control when people make judgments concerning actions' being intentional. An action's being intentional depends, in the above cases, on the answer to the question, "is there any justifying reason not to  $\Phi$ ?".

#### **6.4 CAUSAL DEVIANCE**

In this section, I will show, as in the cases involving skill/luck and unintended side effects, that the moral qualities of the outcome of a behavior in the cases of causal deviance strongly influence people's judgments as to whether that behavior should be considered intentional. I claim that in order to decide whether the effect of an action is intentionally done in the case of causal deviance we need to take account of normative considerations.

Common examples of deviance are two-fold, depending upon what portion of the causal chain gets attention. The first type of deviance, which is called primary deviance<sup>59</sup>, raises a problem about a relatively direct connection of the causal sequence between the motivating mental state that is supposed to cause an action and the bodily movement that is supposed to be the action. Another type of deviance commonly discussed, secondary deviance, locates the problematic event after the bodily movement has occurred. Primary deviance is thought to undermine the very possibility that a bodily movement can count as

---

<sup>59</sup> This is Mele's terminology. See Mele and Moser 1994.



an action<sup>60</sup> in contrast to secondary deviance that apparently undermines the intentional status of an action but allows for the possibility of not-intentional actions. Some varieties of primary deviance, however, have encountered problems with the possibility of counterexamples where events caused and rationalized by mental states do not count as actions.

Davidson (1973: 79) provides an example of primary deviance. There is the case of the rock climber who wants to rid himself of the weight of his partner and believes that loosening his grip on the rope would do that. And his recognition of that so unnerves him that it causes his hand to tremble in such a way that he loosens his hold. Despite the fact that the movement of the climber's hand is caused by the want and the belief, the agent did not, according to Davidson, loosen his hold intentionally.<sup>61</sup> While an appropriate belief/desire pair of intentional attitudes may rationalize the event, some would be reluctant to say that the event of loosening his hold counts as an intentional action as well as an action.

---

<sup>60</sup> Now in the cases of primary deviance our focus is changed into intentional movements, not intentional actions. However my main point works in these cases also.

<sup>61</sup> I will claim that actually the case in question is intentional. I will provide a counterexample to Davidson's view later in this section.

This case counts as a typical case of basic deviance of causal sequence between the motivating mental states and the movement of the climber's hand.

Before we turn to the example of secondary deviance, let us examine why Davidson thinks that the case in question is not intentional. Davidson claims that in this case "he never *chose* to loosen his hold, nor did he do it intentionally" (1973: 79). However, what's the reason for Davidson to say that he did not loosen his hold *intentionally*? Davidson seems to think that it is not intentional because there is no right connection that must obtain between mental antecedents and bodily movement for action to count as intentional.

Davidson says:

Beliefs and desires that would rationalize an action if they caused it in the *right* way - through a course of practical reasoning, as we might try to saying - may cause it in other ways. If so, the action was not performed with the intention that we could have read off from the attitudes that caused it. (1973: 79)

Davidson claims that the belief/desire pair did not cause the action in the right way. Maybe this is enough for him to say that the action in question is not intentional. If this is right, however, Davidson seems to claim that

every primary case of causal deviance is not intentional, and this seems to be incorrect in many respects.

Davidson may claim more than this. As I mentioned in Chapter Five, Davidson, following the Standard Account in a way, gives a necessary condition for action to be intentional; namely the agent must have a reason that rationalize her action. If there is no reason for the agent to  $\Phi$ , then it is not intentional. Davidson may think that the climber's loosening his hand was not intentionally done because the climber did not have any reason that he loosened his grip.

Let us now return to the example of secondary deviance, also discussed by Davidson. Here a man tries to kill someone by shooting him (1973: 78-79)<sup>62</sup>. However, his shot misses his victim by a mile, but makes a herd of pigs stampede, which in turn tramples his target to death. Although the victim's death was caused by an appropriate belief/desire pair, we would not say that the would-be sniper intentionally killed the victim.<sup>63</sup>

---

<sup>62</sup> This is an example of Daniel Bennett's (Bennett 1965).

<sup>63</sup> I claim that this case is also intentional.

In order to deal with the cases of causal deviance I argue that we need to look at the justifying reason.<sup>64</sup> In Davidson's example of trying to kill someone by shooting him, bringing about the effect of "killing someone" was the man's reason for shooting. However, there is a deviance between the shooting and the event of killing. Because of the deviance Davidson is saying that the man did *not* kill the victim *intentionally*. However in the cases of causal deviance, like the lucky actions and unintended side effects, in order to see whether the victim's death was done intentionally we need to ask a

---

<sup>64</sup> In fact, there has been widespread belief that answering the problem of causal deviance adequately is tied directly to the theoretical task of providing necessary and sufficient conditions for the identification of an intentional action. This seems to follow from the fact that the Standard Account may characterize the intentional action in terms of its causal features. And if we take the Davidsonian route that the explanation of action for a reason is a kind of causal explanation, then one can provide a list of the necessary and sufficient conditions for identifying an event as an intentional action on the condition that one can identify the causal conditions required for a chain of events to produce an intentional action. However, we have seen several cases of countering the Standard Account of intentional action.

Others, for example Armstrong (1973), deal with the cases of causal deviance in a way that reasons, if they are to rationalize, must cause action "in the right kind of way." However the effort has turned out to be unsuccessful. Causing an action in the right kind of way is to produce the effect by the right kind of causal route. This solution was also what Davidson followed one time. He tried to solve the problem by saying that the psychological antecedents that bring about action must cause the action "in the right way" if it is to count as intentional movement (1973: 78-79; 1978: 87). Davidson, however, acknowledged that there is some difficulty with attempting to solve the problem of causal deviance by using the locution "in the right way." He said that it not only hardly gives any insight, but actually the search for looking for the meaning of the phrase "in the right way" turns out to be an insurmountable task.

question, "was the fact that death would occur a justifying reason not to perform the shooting?". If the answer is "yes," then it was done intentionally. If "no," it's not intentional. Then, the death may be intentional on the condition that there is a reason not to perform the shooting that results the death.

Let's take a look at Davidson's climber again. Despite the fact that the movement of the climber's hand is caused by his belief/desire pair, it seems not, according to Davidson, to be an intentional bodily movement. Rather, it is a purely accidental bodily movement that happens to match the climber's motivating mental states. However, the fact that if he loosens his grip, then his partner would fall and it would cause him to a death, seems to be enough of a reason, from a third-party perspective, against loosening his grip. If it shows that the action that follows from the belief/desire pair, regardless of causal deviance or not, should not be done for whatever reason, we are bound, if we are rational, not to act on it. This is the case where we have a justifying reason not to loosen his grip. Despite the fact that the causal route was deviant we seem to

want to say, contra Davidson, that the climber's loosening was an *intentional* movement.

If you are not sure about the intuition about the loosening of climber's being intentional, just take a look at Wilson's example of the weightlifter (Wilson 1989: 152). Like the climber's case this is also a case of primary deviance. As Wilson sets up the example, a weightlifter's intention to lift a very heavy weight causes him to become nervous, and that state of agitation provides just the nervous energy necessary for him to succeed in lifting the weight. That is, his accidentally produced state of nervousness is a crucial causal factor in his successful lifting of the weight. If the story ends here, there is no causal deviance. However, suppose the weightlifter should not be nervous, nor intend to get nervous, because studies have shown that getting nervous would sap his strength rather than enhance it. And also suppose that everybody, including the weightlifter, knows the result of the studies. Then the causal route from intention to action was deviant because it was not a route which the lifter intended, nor believed would be successful. He may never have lifted a weight that way before, and he may never do it again. None of these

things prevent the agent's intention to lift a very heavy weight from causing the lifting of the weight. And none of these things undermines, in the slightest, the claim that his lift was an intentional action, which, according to my definition, it surely was.<sup>65</sup> I said that one of the notable features of my criterion for intentional action is that it is not limited to just moral considerations. And this is a merit because it is able to deal with Wilson's weightlifter very easily. The question to be asked in this case is, "is there any justifying reason not to lift the weight by using nervous energy?". And in this case the answer is "yes," making the lifting intentional.

Now imagine Davidson's climber again, but there is only this difference: the climber is holding some baggage instead of his partner. Now, the question to be asked in order to decide whether the movement in question is intentional, is "do we have a justifying reason, from a third-party perspective, for the agent not to loosen his grip?": is the fact that the baggage would fall if he loosens his grip a justifying reason not to loosen his

---

<sup>65</sup> In fact, what the example of the weightlifter suggests, I think, is that the type of causal route from an intention to a bodily movement is simply irrelevant to the movement's being intentional.

grip? Here people's intuitions may depend on what the baggage may have, or maybe something else. If, for example, it has a bomb to be able to kill innocent people, then this gives a enough reason not to loosens his grip and the climber's loosening is said to be intentional. However, we can also imagine lots of cases that there is no justifying reason not to loosen his grip. In these cases the resulting movement is not intentional<sup>66</sup>.

The climber's case shows that people's intuitions may also be different in the case of causal deviance. One merit of my definition, like the cases of lucky actions, is the fact that it explains the different intuition on whether an agent performs a behavior intentionally when the result seems to be due to causal deviance. If I, as opposed to Davidson's intuition, am right, then the climber's example suggests another counterexample to the Standard Account that we saw in Chapter Five; loosening his hold was not done for a reason but it seems to be taken to be intentional.<sup>67</sup>

---

<sup>66</sup> I will call this case C2, while I am calling the original climber's case C1. These two cases will be used in the next section when I am arguing against NCC.

<sup>67</sup> Davidson's second example that borrows from Bennett is not a counterexample to the Standard Account, though. This is the case that shooting was done for a reason but the effect of the shooting is, if I am right, *also intentional*. However, if Davidson's



## 6.5 NCC AND MENTAL CAUSATION DEBATE REVISITED

As I mentioned in Chapter Three, in the case of mechanistic explanations where the why-question is interpreted as a request for a mechanism that we may or may not be able to provide, any response to the why-question also provides information that can adequately answer the how-question. In mental explanation, however, we do not expect the same pattern of interchangeability between why- and how-questions.

In this section I will examine the difference between mental explanation and naturalistic explanation with regard to the why- and how-questions. From this consideration I claim that because of the justificatory factor in dealing with intentional action, it is difficult to see how NCC can be true. I argue that NCC is an error due to confusing a mental explanation with a purely naturalistic explanation between events. I further claim that this insight works nicely in the case of

---

intuition is right, then this case is a counterexample to the Standard Account. With regard to the climber's case, loosening the hold was not done for a reason and it's not, according to Davidson, intentional, therefore the case is not a counterexample to the Standard Account, either.

causal deviance. Finally I argue that NCC is just the result from supposing that there is no gap between explanatory reason and justificatory reason.

Now consider the following questions:

- (1a) Why did Brutus stab Caesar?
- (1b) How did Brutus stab Caesar?

The answer to the first question may be something like "He stabbed Caesar because he wanted to end the tyranny." Let us suppose the following: An instantiation of the property M, Brutus' wanting to end the tyranny, causes an instantiation of the non-mental property N, Brutus' stabbing Caesar. Now Brutus' stabbing is causally explained by his wanting to end the tyranny. However, for someone like Jaegwon Kim, this picture is not enough to give an explanation. He thinks that the instance of M is there, because of M's physical realization base, P. He thinks that we need to provide *how* the event came about by providing a mechanism connecting N and P. However, I argue that this is an error due to confusing a *mental explanation* with a purely *naturalistic explanation* between events.

In order to see the difference between mental and naturalistic explanation consider the following:

- (2a) Why did the house catch on fire?
- (2b) How did the house catch on fire?

In the case where the why-question is interpreted as a request for a mechanism that we may or may not be able to provide, any response to the why-question also provides information that can adequately answer the how-question. To formalize the sentence "A house did catch fire"<sup>68</sup> by using Davidson's apparatus<sup>69</sup>, it will be: "There is an event that is a firing of a house."<sup>70</sup> What the logical paraphrase seems to suggest, is that (2a) and (2b) seek an explanation about the existence of an event, and the natural way to explain this is to present details of the

---

<sup>68</sup> It should be "The house" instead of "A house." It does not, however, make any difference for the purpose of the argument here.

<sup>69</sup> The received view is that "folk-psychological" explanations of action are causal and one reason for accepting it is logical form. Davidson has argued that action-sentences have a logical form that involves quantification over events.

<sup>70</sup> Davidson's contribution in the issue of logical form of an ordinary action sentence like "Brutus stabbed Caesar" is the defense that it has the logical form of an existential generalization. According to Davidson, the logical form of the sentence, "Brutus stabbed Caesar", is an existential generalization:

$(\exists x) (\text{Stabbed} (\text{Caesar}, \text{Brutus}, x))$

This states that there is something that is a stabbing of Caesar by Brutus. Davidson claims that the thing or things that are related to Brutus and Caesar by this sentence - the things over which the sentence quantifies - are events.

event's causation, that is, its coming to be: events come into existence because they are caused by other events.

In mental explanation, however, it is not the case that we expect the same pattern of interchangeability between why- and how-questions. Brutus stabbed Caesar because he wanted to end tyranny. Meanwhile, Brutus' wanting may have been realized by numerous ways. It may be realized by expressing his anger toward tyranny in public speech, or by striking a table hard in front of him, and so on. Kim's asking of M's physical realization base is just to ask something further, namely, "how it is realized."

In fact NCC is just the result of some philosophers, including Kim and Davidson, asking this further thing. Kim may expect some kind of causal mechanism to answer the question (1b), "How did Brutus stab Caesar?," therefore connecting Brutus' stabbing with the realization base of Brutus' wanting to end the tyranny. However as we saw in Chapter Three, the causal relations between mental properties or between the mental and the physical do not depend on causal relations between the properties that realize them. Which microproperties realized Brutus' wanting to end the tyranny depends on

how his wanting was made (by expressing his anger toward tyranny in public speech, or by striking a table hard in front of him). But the effect is indifferent to how it is realized. The answer to the question (1b), "How did Brutus stab Caesar?," does not adduce causal information that explains the occurrence of an event, for we are citing certain actions of Brutus by using a knife or by slashing or some other ways. And the answer to the how-question here does not answer the why-question; Brutus' action of slashing, for instance, does not answer the question "why did Brutus stab Caesar?"

Asking the how-question is just to further presuppose that how-questions and why-questions do not make any difference in mental causation. This holds only in the case of a purely naturalistic explanation between events. Despite the fact that the questions expressed by (2a) and (2b) were equivalent due to receiving the same kind explanation as answers, questions (1a) and (1b) do not receive the same answers. The fact that the answers to the how-questions give some information on the answers to the why-question works only for a purely naturalistic relation between events.

This insight, so far discussed, works nicely in the case of causal deviance. Let's take a look at Davidson's climber case (C1) and the revised case (C2). The two cases are exactly the same except that the climber in C2 is holding some baggage instead of his partner. As you remember, while in C1 the climber should not act on his belief/desire pair because there exists a reason not to act on it, there seems to be no such reason in the latter. While the climber's loosening in C1 was an intentional movement despite being causally deviant, the resulting movement in C2 may not be intentional even though the fact that the movement of the climber's hand is caused by his belief/desire pair; it may be a purely accidental bodily movement that happens to match the climber's intention.

Now the why- and the how-questions are treated differently in two cases. In case C2 the two questions,

- (3a) Why did the climber loosen his grip?
- (3b) How did the climber loosen his grip?

do not make any difference because the answer is found by simply referring to "becoming unnerved." The question (3a) does not require any further explanation than this.

This means we treat the two questions in the case (C2) of not-intentional action the same as in the case we deal with a naturalistic explanation between events. If not-intentional, there is no sense to ask further beyond the why-question, since we expect the same pattern of interchangeability between the why- and how-question, and those two questions do not make any difference. In order for us to ask further, the action in question should be intentional.

The answer to the why-question in C1, which is intentional, however, does not answer the how-question. Answering the question "why did the climber loosen his grip?" simply by referring to his state is not sufficient since it does not capture the *point* that the climber has a normative reason not to loosen his grip. This point, NCC cannot deal with. NCC never deals with reasons not to  $\Phi$ . NCC does not explain why the climber should not have loosened his grip, nor concern normative requirements working in this case (C1) of causal deviance.

## 6.6 CONCLUSION

The problem of mental causation results from some unwarranted metaphysical assumption: the Principle of Nomological Character of Causality (NCC). However, there is little reason to understand causation in the manner required to make NCC work. I mentioned in the first part of this work that the assumption of NCC is responsible for the problem of mental causation. If we reject NCC, we don't need to worry about finding some explanations relating intentional explanations to physical explanations, as far as the intentional explanations are informatively fruitful. Burge puts this point as following:

We determine the nature of causation, and the sort of laws or lawlike generalizations that accompany it, by scrutinizing actual explanations in psychology and ordinary discourse. If there turned out to be no clear sense in which mental events fell under predicates that are uncontroversially physical, then it would seem reasonable to count mental events nonphysical. As far as I can see, there is no reason to be anything but relaxed in the face of this possibility. I see no powerful, clearly articulated reason for worrying about the existence of mind-body causation, or the gaplessness of chains of physical events, if this possibility were realized. What counts in support our belief in mind-body causation is the probity of mentalistic explanations. As long as they are informative and fruitful, we can assume that they are relating genuine events, whatever their metaphysical status. (1992:38-9)



The motivation for the demand for laws in action explanations stems at least in part from the fact that the laws cited in explanations are the laws that subsume events in naturalistic causal relations. By rejecting the idea that causal explanation is causal because it is grounded in natural causal relations, the motivation for requiring laws in explanations disappears. I claim that this is the reason why we need to pay attention to our practice and explanatory strategies.

By rejecting NCC we can in fact arrive at a sustainable, defensible and rewarding account of mental causation. The primacy of explanatory practice over the ontological commitment reverses the usual account according to which causal explanations count as causal if they are grounded in causal relations. However, explanations come first, such that an explanation is causal if we accept it as such. By reinterpreting the notion of causation we regain the causal efficacy of the mental.

The causal efficacy of the mental is not derived from the underlying subvenient properties alone because the causal relations between mental properties or between the mental and the physical do not depend on causal

relations between the properties that realize them. The causal pattern at mental levels, which can occur only in certain circumstances, is not governed by the causal patterns at the lower levels since it cannot be explained by the non-intentional realizing properties which do not consider matters happened in the context or circumstances.

We looked to a theory of intentional action for help in answering the problem of mental causation. I approached the issue of intentional action not by looking into the metaphysics of mind, but by focusing on the role that normative considerations play in our actual explanatory practices in determining whether an action was performed intentionally. I conclude the criteria for intentional action must be wide enough to include the normative perspectives of a third-point of view as well as the psychological perspectives.

## BIBLIOGRAPHY

- Adams, Frederick. 1997. "Cognitive Trying." In G. Holmstron-Hintikka and R. Tuomela, eds. *Contemporaray Action Theory*. Dordrecht: Kluwer.
- Adams, Frederick. 1986. "Intention and Intentional Action: The Simple View." *Mind and Language* 1, pp. 281-301.
- Anscombe, G.E.M. 1957. *Intention*. Oxford: Basil Blackwell.
- Armstrong, David. 1973. "Acting and Trying." *Philosophical Papers* 2, pp. 1-15.
- Audi, Robert. 1986. "Acting for Reasons." *Philosophical Review* 95, pp. 511-546.
- Audi, Robert. 1973. "Intending." *The Journal of Philosophy* 70, pp. 387-403.
- Baker, Lynne Rudder. 2001. "Are Beliefs Brain States." In Anthonie Meijers, ed. *Explaining Beliefs: Lynne Rudder Baker and Her Critics*. Stanford: CSLI.
- Baker, Lynne Rudder. 1999. "What is This Thing Called 'Commonsense Psychology'?" *Philosophical Explorations* 2, pp. 3-19.
- Baker, Lynne Rudder. 1993. "Metaphysics and Mental Causation." In John Heil and Alfred Mele, eds. *Mental Causation*. Oxford: Clarendon.
- Bennett, D. 1965. "Action, Reason and Purpose." *Journal of Philosophy* 62, pp. 85-95.
- Bratman, Michael. 1987. *Intentions, Plans, and Practical Reasons*. Cambridge, Mass: Harvard University Press.
- Burge, Tyler. 1993. "Mind-Body Causation and Explanatory Practice." In John Heil and Alfred Mele, eds. *Mental Causation*. Oxford: Clarendon.

- Burge, Tyler. 1992. "Philosophy of Language and Mind: 1950-1990." *The Philosophical Review* 101, pp. 3-51.
- Burge, Tyler. 1989. "Wherein is Language Social?" in *Reflections on Chomsky*.
- Burge, Tyler. 1989a. "Individuation and Causation in Psychology." *Pacific Philosophical Quarterly* 70, pp. 303- 322.
- Burge, Tyler. 1988. "Individualism and Self-Knowledge." *Journal of Philosophy* 85, pp. 649-663. Burge, Tyler. 1979.
- Burge, Tyler. 1979. "Individualism and the Mental." *Midwest Studies in Philosophy* 4, pp. 73-121.
- Butler, Ronald. 1978. "Report on *Analysis* 'Problem' 16." *Analysis* 38, pp. 113-118.
- Churchland, Patricia. 1986. *Neurophilosophy: Toward a Unified Science of the Mind-Brain*. Cambridge: MIT Press.
- Davidson, Donald. 2004. *Problems of Rationality*. Oxford: Clarendon.
- Davidson, Donald. 2001. *Subjective, Intersubjective, Objective*. Oxford: Clarendon.
- Davidson, Donald. 2001a. *Essays on Actions and Events*. 2<sup>nd</sup> ed. Oxford: Clarendon.
- Davidson, Donald. 1993. "Thinking Causes." In John Heil and Alfred Mele, eds. *Mental Causation*. Oxford: Clarendon.
- Davidson, Donald. 1991. "Epistemology Externalized." *Dialectic* 45, pp. 191-210. Page references to the reprint in *Subjective, Intersubjective, Objective*. Oxford: Clarendon, 2001.
- Davidson, Donald. 1989. "What is Present to the Mind." In J. Brandl and W. Gombocz, eds. *The Mind of Donald Davidson*. Amsterdam: Rodopi. Page references to the

reprint in *Subjective, Intersubjective, Objective*.  
New York: Clarendon, 2001.

Davidson, Donald. 1988. "The Myth of the Subjective."  
In M. Benedikt and R. Berger, eds. *Bewusstsein,  
Sprache und Die Kunst*. Vienna: Edition S. Verlag Der  
Oesterreichischen Staatsdruckerei. Page references  
to the reprint in *Subjective, Intersubjective,  
Objective*. New York: Clarendon, 2001.

Davidson, Donald. 1987. "Knowing One's Own Mind."  
*Proceedings and Addresses of the American  
Philosophical Association* 60, pp. 441-458. Page  
references to the reprint in *Subjective,  
intersubjective, Objective*. Oxford: Clarendon, 2001.

Davidson, Donald. 1984. *Inquiries into Truth and  
Interpretation*. Oxford: Clarendon.

Davidson, Donald. 1980. *Essays on Actions and Events*.  
Oxford: Clarendon.

Davidson, Donald. 1978. "Intending." In Yirmiyahu Yovel,  
ed. *Philosophy of History and Action*. Holland: D.  
Reidel Publishing Company. Page references to the  
reprint in *Essays on Actions and Events*. Oxford:  
Clarendon, 1980.

Davidson, Donald. 1973. "Freedom to Act." In Ted  
Honderich, ed. *Essays on Freedom of Action*. London:  
Routledge. Page references to the reprint in *Essays  
on Actions and Events*. Oxford: Clarendon, 1980.

Davidson, Donald. 1971. "Agency." In R. Binkley, R.  
Bronaugh, and A. Marras, eds. *Agent, Action, and  
Reason*. Toronto: University of Toronto Press. Page  
references to the reprint in *Essays on Actions and  
Events*. Oxford: Clarendon, 1980.

Davidson, Donald. 1970. "Mental Events." In Lawrence  
Foster and J.W. Swanson, eds., *Experience and  
Theory*. The University of Massachusetts Press.  
Page references to the reprint in *Essays on Actions  
and Events*. Oxford: Clarendon, 1980.

- Davidson, Donald. 1969. "How is Weakness of the Will Possible?" In Joel Feinberg, ed. *Moral Concepts*. Page references to the reprint in *Essays on Actions and Events*. Oxford: Clarendon, 1980.
- Davidson, Donald. 1967. "Causal Relations." *Journal of Philosophy* 64, pp. 691-703. Page references to the reprint in *Essays on Actions and Events*. Oxford: Clarendon, 1980.
- Davidson, Donald. 1963. "Actions, Reasons, and Causes." *Journal of Philosophy* 60, pp. 685-700. Page references to the reprint in *Essays on Actions and Events*. Oxford: Clarendon, 1980.
- Dennett, Daniel C. 1987. *The Intentional Stance*. Bradford: MIT press.
- Dupre, John. 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge: Harvard University Press.
- Fodor, Jerry. 1987. *Psychosemantics*. Bradford: MIT Press.
- Ginet, Carl. 1990. *On Action*. Cambridge, Mass.: Cambridge University Press.
- Gordon, Robert. 1986. "Folk Psychology as Simulation." *Mind and Language* 1, pp. 158-171.
- Harman, Gilbert. 1976. "Practical Reasoning." *Review of Metaphysics* 79, pp. 431-63.
- Heil, John, and Alfred Mele, eds. 1993. *Mental Causation*. Oxford: Clarendon.
- Honderich, Ted. 1982. "The Argument for Anomalous Monism." *Analysis* 42, pp. 59-64.
- Kenny, A. 1963. *Action, Emotion and Will*. London: Routledge and Kegan Paul.
- Kim, Jaegwon. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, Mass.: MIT Press.

- Kim, Jaegwon. 1995. "Mental Causation: What? Me Worry?" *Philosophical Issues* 6, pp 123-51.
- Kim, Jaegwon. 1993. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press.
- Kim, Jaegwon. 1993a. "Can Supervenience and 'Non-Strict' Laws Save Anomalous Monism?" In John Heil and Alfred Mele, eds. *Mental Causation*. Oxford: Clarendon.
- Kim, Jaegwon. 1993b. "The Nonreductivist's Troubles with Mental Causation." In John Heil and Alfred Mele, eds. *Mental Causation*. Oxford: Clarendon. Page references to the reprint in Jaegwon Kim, *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press, 1993.
- Kim, Jaegwon. 1989. "The Myth of Nonreductive Materialism." *Proceedings and Addresses of the American Philosophical Association* 63, pp. 31-47.
- Kim, Jaegwon. 1989a. "Mechanism, Purpose, and Explanatory Exclusion." *Philosophical Perspectives* 3, pp. 77-108. Page references to the reprint in Jaegwon Kim, *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press, 1993.
- Kim, Jaegwon. 1984. "Epiphenomenal and Supervenient Causation." In *Midwest Studies in Philosophy* 9, pp. 257-70.
- Kim, Jaegwon. 1982. "Psychophysical Supervenience." *Philosophical Studies* 41, pp. 51-70. Page references to the reprint in Jaegwon Kim, *Supervenience and Mind: Selected Philosophical Essays*. Cambridge: Cambridge University Press, 1993.
- Knobe, Joshua. 2004. "Intention, Intentional Action and Moral Considerations." *Analysis* 64, pp. 181-187.
- Knobe, Joshua. 2003. "Intentional Action in Folk Psychology: An Experimental Investigation." *Philosophical Psychology* 16, pp. 309-24.

- Knobe, Joshua. 2003a. "Intentional Action and Side Effects in Ordinary Language." *Analysis*, 63, pp. 190-194.
- Kripke, Saul. 1976. "A Puzzle About Belief." In Avishai Margalit, ed. *Meaning and Use*. Dordrecht: Reidel.
- MacDonald, Cynthia. 1989. *Mind-Body Identity Theories*. New York, NY: Routledge.
- Malle, B.F. and Knobe, Joshua. 1997 "The Folk Concept of Intentional Action." *Journal of Experimental Social Psychology*, 33, pp. 101-121.
- McCann, Hugh. 1986. "Rationality and the Range of Intention." *Midwest Studies in Philosophy* 10, pp. 191-211.
- McLaughlin, Brian. 1993. "On Davidson's Response to the Charge of Epiphenomenalism." In John Heil and Alfred Mele, eds. *Mental Causation*. Oxford: Clarendon.
- McLaughlin, Brian. 1989. "Type Epiphenomenalism, Type Dualism, and the Causal Priority of the Physical." In *Philosophical Perspectives 3: Philosophy of Mind and Action Theory*. California: Ridgeview publishing Co.
- Melden, A. I. 1961. *Free Action*. London: Routledge and Kegan Paul.
- Mele, Alfred R. 2001. "Acting Intentionally: Probing Folk Notions." In B.F. Malle, L. J. Moses, and D. A Baldwin eds. *Intentions and Intentionality*. Cambridge: MIT Press.
- Mele, Alfred R. and Paul K. Moser. 1994. "Intentional Action." *Nous* 28, pp. 39-68. Page references to the reprint in Alfred R. Mele, ed. 1997. *The Philosophy of Action*. New York: Oxford.
- Mele, Alfred R. and Sverdlik, S. 1996. "Intention, Intentional Action, and Moral Responsibility." *Philosophical Studies* 82, pp. 265-287.



- Peacocke, Christopher. 1985. "Intention and Akrasia." In M. Hintikka and B. Vermazen, eds. *Essays on Davidson*.
- Quine, W.V.O. 1951. "Two Dogmas of Empiricism." *Philosophical Review* 60, pp. 20-43.
- Smith, Michael. 1995. *The Moral Problem*. Cambridge, Mass.: Blackwell.
- Stich, Stephen. 1983. *From Folk Psychology to Cognitive Science: The Case Against Belief*. Bradford: MIT Press.
- van Gulick, Robert. 1993. "Who's In Charge Here? And Who's Doing All the Work?" In John Heil and Alfred Mele, eds. *Mental Causation*. Oxford: Clarendon.
- Vermazen, Bruce and Merrill Hintikka, eds. 1985. *Essays on Davidson: Actions and Events*. Oxford: Oxford University Press.
- Wilson, George. 1989. *The Intentionality of Human Action*. Stanford: Stanford University Press.