

Contributions to a Physicalistic Theory of Action

by

Steven H. Simon

A.B., Cornell University (1981)

Submitted to the Department of Linguistics and Philosophy
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

Massachusetts Institute of Technology

June 2002

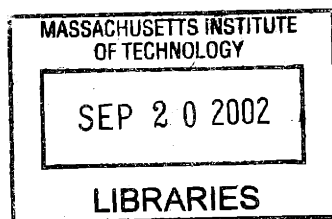
© 2002 Steven H. Simon. All rights reserved.

The author hereby grants to MIT permission to
reproduce and publicly distribute paper and electronic
copies of this thesis document in whole or in part.

Signature of Author
Department of Linguistics and Philosophy
January 30, 2002

Certified by.....
Robert Stalnaker
Laurance S. Rockefeller Professor of Philosophy
Dissertation Supervisor

Accepted by.....
Vann McGee
Professor of Philosophy
Chair of the Committee on Graduate Students



ARCHIVES

Contributions to a Physicalistic Theory of Action

by

Steven H. Simon

Submitted to the Department of Linguistics and Philosophy
on January 30, 2002 in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Abstract

My project of giving a general physicalistic reduction of action contrasts with Donald Davidson's view that only individual actions can be explained in physicalistic terms. The main reason for his view is that he thinks the problem of internal causal deviance is insoluble. In the first chapter, I reconstruct the theory of action Davidson develops in *Essays and Events* and extend the theory to solve the deviance problem. The idea of the solution is that action requires "modulated movement," an ongoing process of monitoring and modulating the movements in which actions consist.

In the second chapter, I develop the theory of modulated movement in more detail and argue that it can explain a number of cases of defective agency. I defend my contention that the analysis of modulated movement solves the deviance problem against several objections. In doing so, one of the main points I argue is that "ballistic movements," movements the agent cannot modify, cannot be actions.

The psychological states in terms of which I analyze modulated movement are belief and desire, and in the third chapter I develop a reductive physicalistic account of a component of belief, indication. I start with a theory of indication that Robert Stalnaker presents in *Inquiry*, and develop the theory to cope with some problems for it that I identify. In the second part of the chapter, I extend the theory to explain cases of indication in which indicators are combined so that together they indicate propositions more specific or precise than any of the propositions they indicate alone, thus reducing complex cases of indication to simpler ones.

Thesis Supervisor: Robert Stalnaker

Title: Laurance S. Rockefeller Professor of Philosophy

Acknowledgment

Many people have helped in the writing of this dissertation. Alex Byrne and Ned Hall provided many helpful comments. John Perry suggested the project and helped me get started. Michael Bratman taught me a great deal about action theory, and shaped my thinking on the topic. My biggest debt is to my thesis supervisor, Bob Stalnaker, whose extensive and insightful comments vastly improved the dissertation. Thanks also go to my friends and family.

Contents

Abstract	3
Acknowledgement	4
Contents	5
Introduction	7
Chapter 1	
Causal deviance and Davidson's theory of action	10
Chapter 2	
Modulated movement	49
Chapter 3	
Indication	94
Endnotes	133
Bibliography	139

Introduction

This project began as an attempt to give a comprehensive theory of action. The goal was to explain action as a widespread phenomenon, encompassing not just the intelligent, rational behavior of human beings, but the goal-directed behavior of much simpler organisms, and the systematic, useful operation of simple mechanisms. What would unify the theory would be the form of the explanations. I wanted to give a general physicalistic reduction of these disparate phenomena – to explain the behavior of all these kinds of things in terms of the spatiotemporal properties of, and the spatiotemporal and causal relations among, the parts that make them up. I hoped that by explaining these phenomena in this common framework, I could bring out their similarities as well as their differences, and could show how the more complex phenomena could be produced by making incremental changes to the simpler ones. A comprehensive theory of action should also explain such phenomena as weakness of will, coercion, manipulation, static action and collective action, and should reconcile the first-person and third-person views of action. I hoped to use the framework to explain these things as well.

However, the project as I originally envisioned it turned out to be much bigger and much more complicated than I had expected. I have not produced the comprehensive theory I had in mind, but, instead, parts of it. To see how these parts fit into the larger project, it will be useful to say a bit more about the comprehensive but non-actual theory. In its broad outlines, it would be a causal theory of action. As such, it would explain actions as events with certain causes. Like most causal theories of action, the theory would hold that actions are events caused by psychological states – reasons – of the agent that brings them about. It would be the job of a subsidiary theory, a theory of rationalization, to explain what reasons are – that is, to say what simpler psychological states constitute them, and how those states are related to one another and to the action. To make the theory of rationalization part of a reductive physicalistic theory of action, there would also be

theories providing physicalistic reductions of the psychological states adduced in the theory of rationalization.

My dissertation consists of three chapters. My point of departure in the first is the causal theory of action of Donald Davidson. Davidson's theory bears on my project in two important ways. First, though causal theories of action go back to Hume and the early compatibilists, it is probably fair to say that Davidson's theory, more than any other, has revived interest in recent years in causal theories of action, and has set the terms of discussion of such theories. So Davidson's theory is worthy of consideration in its own right, and understanding it and what problems it faces is likely to illuminate many other theories of action. Second, though Davidson never abandoned the view that the only relation that need hold between a reason and an event to make the event an action is the causal relation, he came to believe that not just any causal chain from a reason to the event would do. For if the causal chain is *deviant*, then the agent's responsibility for the event will seem diminished or absent. But Davidson thinks it is impossible to say in general, in physicalistic terms, what the difference is between normal and deviant causal chains from reasons to actions. If he is right, then no physicalistic reduction of action can be given.

The problem of internal causal deviance, deviance in the causal chains from reasons to potential actions, is, as I understand it, the main argument Davidson has against the possibility of giving a physicalistic reduction of action. So showing, contra Davidson, that the problem is not insurmountable will go some distance towards making my project more plausible. My aim in the first chapter is to set up the problem of causal deviance in the context of Davidson's project, sketch the idea of a solution, and explain why Davidson might have thought the problem insurmountable, and my aim in the second chapter is to develop the solution in more detail and defend it against some problems that it faces. The solution has the form of an analysis of a kind of movement, modulated movement. When an agent carries out a modulated movement, she

receives feedback about the movement as she carries it out, and responds to the feedback by modifying the movement as necessary to achieve her goal in performing it. It is my contention that all actions involve modulated movement, and whether a movement is caused normally can be determined by considering whether it is properly modulated. My analysis of modulated movement is to a large extent a theory of rationalization. It is based on Davidson's theory of rationalization, extended to allow for the agent to receive and respond to feedback about the action.

As I have mentioned, a reductive physicalistic theory of action requires reductive physicalistic theories of the psychological states used to rationalize action. In defending my solution to the problem of causal deviance in the second chapter, I sketch reductive physicalistic analyses of belief and desire, the psychological states in terms of which Davidson explains reasons. My aim in the third chapter is to develop in more detail the reductive physicalistic theory of one component of belief, indication. One of the main things I try to do is show how complex or distributed indicators can be built out of other, simpler indicators, and, as a result, can indicate propositions more precise than any of their constituents can. In this way, I hope to go some distance towards giving a reductive physicalistic theory of belief, and to suggest what a detailed physicalistic reduction of the psychological states that figure in the etiology of action might look like.

Chapter 1

Causal Deviance and Davidson's Theory of Action

1.1 Conceptual dualism and causal deviance

1.1.1 Reductive physicalism, anti-physicalism, and anti-reductive physicalism

Can we explain human action in the same terms we use to explain things that merely happen? On one side, our subjective experience of the world, the apparent complexity of distinctively human phenomena, and many commonsense intuitions suggest there is an unbridgeable gap between distinctively human phenomena and the rest of the world. On the other side, much recent scientific evidence and many commonsense intuitions suggest we should take a *gradualistic* view of the world, and regard distinctively human phenomena as incremental outgrowths of simpler objects and phenomena. The issue at stake is whether *reductive physicalism* about action is true - whether the world is both ontologically and explanatorily unified at the physical level, so that we can explain everything there is using a single, physicalistic explanatory framework. We can make the difference between the two views more vivid by considering what each implies about the possibility of building something capable of human action. The opponent of reductive physicalism believes that the only way, even in principle, in which we can build something capable of human action is by copying molecule-by-molecule something else that is capable of human action. The friend of reductive physicalism, by contrast, believes that we can discover a body of general principles that would enable us to build things capable of human action, or at least that would enable us to understand how to do so.

Donald Davidson's theory of action is a version of *anti-reductive physicalism*. As a physicalistic theory, it assumes that everything that exists is ultimately physical, and implies that the world is ontologically all of a piece. In the particular case of events that are actions, it assumes that they have a level of description at which they can be characterized in

physicalistically acceptable terms - in terms of simple physical objects, such as molecules or atoms, standing to one another in spatiotemporal and causal relations, and relations of similarity.

We can contrast physicalistic theories with theories that are *anti-physicalistic* such as Roderick Chisholm's early theory of agent-causation.¹ Unlike physicalistic theories, anti-physicalistic theories imply that events that are actions do *not* have a level of description at which they can be characterized in strictly physical terms. For instance, in the case of Chisholm's early theory, an action is a matter of a person *agent-causing* something, and, on the theory, there is no level of description of a person agent-causing something at which it can be characterized in strictly physical terms: agent causation is ontologically distinct from, and cannot be reduced to, the physicalistically acceptable relation of *event* causation, the kind of causation that, on the theory, figures in explanations of all events except for actions. Thus anti-physicalistic theories, unlike physicalistic theories, imply that the *phenomenon* of action cannot be reduced to simpler phenomena, and so that ontological dualism is true.

But though, as a physicalist, Davidson believes the *phenomenon* of action *can* be reduced to simpler phenomena, unlike the reductive physicalist, Davidson believes the *concept* of action *cannot* be reduced to simpler concepts. Thus both the reductive physicalist and the anti-reductive physicalist, like Davidson, believe that actions have a level of description at which they can be characterized in physical terms, but they disagree about the relation of the physical level of description to the ordinary level of description: the reductive physicalist believes it is possible to give a general, systematic account of the relation between the two levels of description, while the anti-reductive physicalist believes it is impossible to give such an account. Davidson's view of the relation between the ordinary and physical levels of description is encapsulated in his well-known thesis of *anomalous monism*, the thesis that every *individual* mental event is identical to an *individual* physical event, but there are no laws that link *types* of mental events to *types* of physical events.

Explaining the irreducibility of being intentional as a special case of anomalous monism has the following consequences for the relation of actions to the rest of the world. Action supervenes on physical objects having physical properties and standing in physical relations – on molecules having certain shapes and causal powers, and standing in certain spatial and causal relations to other molecules. So *ontologically* – with respect to the kinds of things there are, the properties they have, and the relations among them – the world is all of a piece. On the other hand, it is impossible to give a general, systematic explanation of action in such simple, reductive terms, and so *conceptually* – with respect to how we *conceive* of the kinds of things there are, their properties, and their interrelations – the world falls into two parts: it is a dichotomy of the non-agential, which we can conceive of in simple physical terms, and the agential, which we can conceive of only in terms of notions that we apply in the first instance to people and their intelligent behavior. Thus I view the disagreement between the reductive physicalist and the anti-reductive physicalist as a disagreement about the nature of the concepts in terms of which we can explain or understand the things there are, not about what kinds of things there are. Since on Davidson's view the world is unified but we cannot conceive of it that way, I think we can see why he characterizes his view of the relation of agents and actions to the rest of the world and of the relation of the mental to the physical, as "ontological monism coupled with conceptual dualism."²

In considering the relation of Davidson's theory of action to reductive physicalism, on the one hand, and to anti-physicalism, on the other, I think it is instructive to consider the historical context of his theory. In his first essay on action, "Action, Reasons, and Causes,"³ he identifies his targets as philosophers like Ryle, Anscombe and Melden, who denied that explanations of action in terms of reasons of the agent are causal explanations⁴ - and so that they can be assimilated to explanations in the natural sciences and to our ordinary explanations of non-agential phenomena. As part of a large-scale project of giving a physicalistic account of the world that shows how

people can be part of the natural order, Davidson's theory of action, though it is anti-reductive, can still be seen as a significant step forward when considered against the background of the theories that were his targets. I think the anomalous monism of his theory of action is attractive as a compromise position: it explains the world as unitary, but in a way that preserves the intuition that human beings and their behavior are too complex to explain in the same way in which we explain simple inanimate objects and their behavior. But I also think there is compelling scientific evidence, much of it fairly recent, that suggests, in contradiction of Davidson's theory, that we *can* give a physicalistic account of human behavior. So, it seems to me, in his acceptance of conceptual dualism Davidson does not go as far as recent developments in the sciences - particularly evolutionary biology, biochemistry, and animal and human physiology - suggest we can go in assimilating people to the natural order

1.1.2 Causal analyses and physicalistic reductions

The anti-reductionism of Davidson's theory of action is striking because the theory is physicalistic. It is also striking because, as Davidson seems to have thought of his theory initially, it was consistent with reductive physicalism, and not explicitly anti-reductive. At the outset, Davidson's project seems to be to give a *causal analysis* of action – an analysis that has the following form:

- (D) For every x , x is an action of an agent a if and only if there is some event y such that y involves a 's instantiation of some property ϕ , and y causes x .

I will say that one who endorses a definition of action that has this form endorses *causalism about action*. The form of causalism suggested by Davidson's original theory is one on which ϕ is

elucidated in terms of psychological states of the agent. Roughly, the view is that an action is a movement of an agent's body caused by her having a reason to make the movement, where a person has a reason to do something if and only if she has a desire that she do it and a belief that she has a way to do it. (I will consider the view in more detail in section 1.2.1.)

What is the relation between causalism about action and reductive physicalism about action?⁵ For a causal analysis of action to be a reductive physicalistic theory, ϕ in (D) must be elucidated in strictly physical terms. I will say that a definition like this provides a *physicalistic reduction* of action. Causalism about action is neither necessary or sufficient for a physicalistic reduction of action. It is not necessary because one might conceivably give a general account of action in physicalistic terms – which is all that is required for a physicalistic reduction of action – without claiming that actions are events with certain causes. It is not sufficient because one might accept a definition of action that has the form of (D) but deny that ϕ can be elucidated in strictly physicalistic terms. For instance, an anti-physicalist might hold that an event is an action if and only if it is caused by an event that includes a certain kind of immaterial substance. But a causal analysis of action provides a natural point of departure for a physicalistic reduction of action. The strategy I propose for giving a physicalistic reduction of action is to start with the causal analysis of action in terms of psychological states of the agent that I described. Then the physicalistic reduction is completed by giving a physicalistic reduction of those psychological states. (I will discuss how this might be done in sections 2.4.1 and 2.4.3 of chapter 2, and in chapter 3.)

Davidson never claims that his theory of action, as he originally thought of it, is consistent with reductive physicalism about action. But in the absence of reasons to reject the possibility of a physicalistic description of ϕ in (D), a theory that provides a causal analysis of action is consistent with reductive physicalism about action. However, Davidson later came to think that no causal analysis of action can be given. I will say that one who believes action can be explained without adducing any relations besides causal relations (or spatiotemporal relations or

relations of similarity) endorses a *causal theory of action*. One can endorse a causal theory of action without endorsing causalism about action. For instance, one might claim that only causal relations are needed to explain action, but deny that just any causal path to an event from a reason to bring it about will suffice for the event to be an action. This is the view that Davidson eventually came to hold. Thus there seems to be a tension not only between Davidson's eventual view of action and reductive physicalism about action, but also between how Davidson initially envisioned his project of explaining action and how the project actually turned out.

1.1.3 Davidson's objections to reductive physicalism

What led Davidson to think a physicalistic reduction of action is impossible? I have already mentioned one reason. Davidson assumes anomalous monism is true, so there are no psycho-physical laws, so types of mental events cannot be reduced to types of physical events, and the mentalistic notions in terms of which action are understood cannot be expressed in physical or behavioral terms.⁶ This particular reason for rejecting reductive physicalism, however, is an assumption that Davidson made - a datum of his theory of the relation between the mental and the physical⁷ - and I will set it aside. Instead, I want to consider reasons he gives that would be reasons for rejecting reductive physicalism even if we assume there *can* be psycho-physical laws that would, in the absence of those reasons, make it possible to give a systematic reduction of mental events to physical events. In the case of these reasons, anti-reductive physicalism was not an assumption Davidson made at the outset, but the result of a concession to problems he encountered as he considered whether a causal analysis of action is possible. One such reason is that, given the fact that persons typically have many different desires whose satisfaction would require doing different, inconsistent things, one cannot give a substantive explanation of why the agent satisfies some at the expense of others.⁸ A strategy the reductive physicalist might follow to solve this problem is to postulate that in addition to beliefs and desires, the etiology of action also

includes *intentions* whose function is to settle on specific desires to satisfy,⁹ and then to give a physicalistic reduction of intention.¹⁰

But "the more general difficulty"¹¹ that Davidson sees is the *problem of causal deviance*. We can distinguish two kinds of causal deviance, *external* and *internal*. The following example illustrates external causal deviance.¹² Smith has a reason to kill Jones, and she tries to do so by shooting him. But she misses, the shot causes a herd of wild pigs to stampede, and the stampede kills Jones. So Smith kills Jones as the result of a *deviant* rather than *normal* causal chain. In a case like this, Smith seems to be less responsible for killing Jones than if her shot had been on target. We can illustrate internal causal deviance with another example.¹³ Smith and Jones are mountain climbers, and Smith holds a rope that supports Jones. Realizing she can rid herself of the weight and danger of supporting Jones by letting go of the rope, Smith forms a reason to do so. But her awareness of her reason makes her panic, and her panic causes her to loosen her grip and let go of the rope. Smith seems less culpable in a case like this, in which her reason causes her movement via a deviant causal chain, than if her reason caused her movement via a normal causal chain, and the movement were more fully intentional.

Though Davidson believes the external problem is serious, it is the internal problem that, he says, "seems...insurmountable."¹⁴ This is the main reason of which I am aware for Davidson's rejection of reductive physicalism about action: he thinks the possibility of internal causal deviance precludes giving sufficient causal conditions for action, and so giving a causal analysis of action. If this is right, we cannot follow the strategy of giving a physicalistic reduction of action by first giving a causal analysis of action, and then expressing the analysis in physicalistic terms. It follows that we must either give up on the project of giving a physicalistic reduction of action or abandon the strategy of first giving a causal analysis of action. But if we give up on the project of giving a physicalistic reduction of action, we give up on answering such questions as how we could systematically build an agent – although there is considerable

scientific evidence that a gradualistic conception of the world is true, which suggests, in turn, that it should be possible to give a physicalistic reduction of rational human behavior. Nor is the alternative of abandoning causalism about action any more attractive – for one thing, it is not at all apparent what might take its place. The stakes are clear. If Davidson is right about the insolubility of the problem of internal causal deviance, then we must either give up on the project of giving a physicalistic reduction of action or find an alternative to causalism. But if Davidson is wrong and the problem of internal causal deviance can be solved, then it poses no obstacle to giving a physicalistic reduction of action.

In section 1.2, I will describe Davidson's theory of action in detail, and will consider how the problem of causal deviance bears on the theory. I will argue that his theory is better off if there is a solution to the problem of causal deviance than if there is not. In section 1.3, I will describe what I think is a promising strategy for dealing with the problem of internal causal deviance in a way consistent with reductive physicalism. I will focus on the problem of internal causal deviance, and for the most part will set aside the problem of external causal deviance, since it is the internal problem that Davidson thinks insoluble. Finally, in section 1.4, I will consider why Davidson might think that the problem of internal causal deviance is insoluble. My ultimate conclusion will be that what most likely leads to his pessimism is a mistaken general view of action and intentional movement.

1.2 Davidson's theory of action

To characterize the effect on Davidson's theory of action of his view that the problem of internal causal deviance is insoluble, I will first describe the theory, and will then say how Davidson responded to the problem of causal deviance. Davidson does not state the theory or the response explicitly, at least not as I will formulate them. However, my formulations are based on various claims and analyses that he makes in different essays in *Actions and Events*.¹⁵ My

contribution is to collect the claims and analyses and put them together to reconstruct his theory of action, filling in a few holes as needed. Thus the form in which I present the theory is mine, but the content, with a few exceptions, is Davidson's – provided I have understood correctly his claims and analyses, and how to put them together. (As I present the theory, I will make explicit what is due to Davidson, and what I have added.)

The theory is suggested by things Davidson says in “Actions, Reasons, and Causes,” and in a later essay, “Agency.”¹⁶ The response to causal deviance is suggested by things he says in the essay “Freedom to Act,”¹⁷ written after the other two essays. As we will see, the crucial effect of his response is to change his theory from one that provides a causal analysis of action, or at least holds out the possibility of doing so, to one inconsistent with a causal analysis. My discussion of his theory of action is divided into two parts. In section 1.2.1 I describe the basic features of his theory, and in section 1.2.2 I describe a component of his theory of action, his theory of agency. I discuss his response to internal causal deviance in section 1.2.3.

1.2.1 The basic theory

A widespread view is that the problem of explicating the difference between actions and passive happenings is the most fundamental problem of action. On Davidson's view, the difference between them can be made out in terms of the notion of *being intentional*. There are many things we *do*, but only some are *intentional*; these are our *actions*. Davidson, however, rejects the view that events are intentional *tout court* for the view that events are intentional *under a description*. (I will discuss why in section 1.2.2.1 when I consider his theory of action descriptions.) We can state his definition of action as:

- (A) An event *e* is an action of a person *p* if and only if *e* is intentional under a description for *p*.¹⁸

The notion of being intentional under a description is explained in simpler terms as follows.

- (I) An event e is intentional under a description q for a person p if and only if¹⁹
- there exists a reason r to make it the case that q such that:
- (a) p has r ,²⁰
 - (b) p 's having r causes e ,²¹ and
 - (c) e is a movement of p 's body.²²

Note that as I have formulated clause (b), I have attributed to Davidson a notion of event that is used in an artificially broad way, to include things that are more abstract than events and that might more naturally be called "states" or "states of affairs."²³ Note also that, as I will understand it, clause (b) has the effect of excluding as actions cases in which, say, a person raises her left hand, but does so by picking it up with her right hand. For in that case, what seems more natural to say is that her left hand went up because she picked it up with her right hand, not that it went up because she wanted it to. (As we will see, though movements like the agent's raising her left hand in the example cannot be intentional on our definition, and so cannot be actions, they can still be intentional in a derivative sense, as effects of actions.)

We define having a reason as follows:

- (R) A person p has a reason to make it the case that q if and only if:
- (a) p desires that q , and
 - (b) p believes that she has a way to make it the case that q .²⁴

Belief and desire are to be understood quite generally, so that belief includes such other information-bearing states as knowing and perceiving, and desire includes such other "pro attitudes" as wanting and regarding as morally or aesthetically desirable.²⁵ Note that (R) excludes cases in which, say, a person moves her big toe when she is trying to move her little toe. I think we should exclude them since, pretheoretically, it seems right to say that the person has a reason to move her little toe, but not that she has a reason to move her big toe.

To illustrate the definitions, consider an event that has the description, "Smith moves her finger." The event, Smith's moving her finger, is Smith's action since, we suppose, it is intentional for Smith under the description, "Smith moves her finger." The event is intentional for her under the description since, we suppose, there is a reason *r* to make it the case that Smith moves her finger such that (a) Smith has *r*, (b) Smith's having *r* causes Smith to move her finger, and (c) Smith's moving her finger is a movement of Smith's body. Finally, Smith has a reason to make it the case that she moves her finger since, we suppose, (a) Smith desires that she move her finger, and (b) Smith believes that she has a way to make it the case that she moves her finger.

A question that arises, since clause (c) of (I) requires that an event be a movement to be intentional, is how Davidson's theory copes with mental actions, intentional mental events that are not movements of an agent's body, such as remembering, deciding or counting to oneself. At one point Davidson suggests the notion of a bodily movement should be construed broadly enough to cover these cases,²⁶ but later he suggests they should be set aside.²⁷ Since the broad construal of movement seems to be artificially broad in light of the ordinary notion, I will take his second suggestion, and will set mental actions aside.

What is important for our purposes about the analysis is that it suggests that when we say that an event is intentional for a person, what we basically mean is that a reason of the person was the cause of the event.²⁸ If we can explain the relation of being intentional - a relation that holds between the action and the agent - in terms of strictly causal relations between the action and the

agent's having reasons, then we can explain actions as events produced by a certain sort of event involving an agent. In our terminology, the explanation provides a causal analysis of action. If we can explain action in this way, then we need not rely on any relations different in kind from those we use to explain all other, non-agential phenomena, and can assimilate ordinary explanations of action to ordinary commonsense and scientific causal explanations, the sorts of explanations we use to explain things that happen to us and events involving only inanimate objects. Thus Davidson's theory, as I have so far reconstructed it, solves the problem of explaining the difference between actions and happenings in a way consistent with reductive physicalism, and holds open the possibility of achieving a physicalistic reduction of action by going on to give a physicalistic reduction of belief and desire.

1.2.2 Davidson's treatment of the problem of agency

I turn now to the problem of agency. Davidson construes it as the problem of explaining the *agent relation*, the relation between a person and an event that makes the event her action.²⁹ I will state Davidson's definition of *being an agent* accordingly:

(AG) A person *p* is the agent of an event *e* if and only if *e* is an action of *p*.

Thus for Davidson the notion of agency, like the notion of action, is tied to the notion of being intentional.³⁰ I will extend the notions of causal analysis and physicalistic reduction, and will say that (AG) provides a causal analysis/physicalistic reduction of *being an agent* if it is part of a theory that includes a causal analysis/physicalistic reduction of action.

Davidson's treatment of the problem of agency is of particular relevance to our concerns for two reasons. First, the problem of causal deviance poses a threat to the simplicity and elegance of Davidson's theory of agency, a theory that, as he says, provides "a vast simplification

of the problem of agency."³¹ As we will see, the simplification - which I will describe in section 1.2.2.1 when I consider Davidson's theory of *action descriptions* - enables us to give a causal analysis of agency by explaining a single instance of a single kind of relation between an agent and a single event, her action, in causal terms. As we will also see, an explanation of this sort follows naturally from Davidson's theory of action. Without the simplification, the problem of giving a causal analysis of action would be much harder, since when an agent acts there would be not a single event to which she stands as agent but many, and her relation to some of the events could be different from her relations to others, so that there could be multiple kinds of relations that would have to be explained. If the problem of external causal deviance can be solved, the simplification can still be achieved, so that when an agent acts there is a single event to which she stands as agent. But if the problem of internal causal deviance is insoluble, the simplification is the end of the road for our attempt at a causal analysis of agency, since no causal analysis of the relation between the agent and the single event can be given.

The second reason for considering Davidson's treatment of agency is that, while Davidson says he thinks the problem of internal causal deviance is insurmountable, he does not say why, and I think his treatment of the problem of agency suggests, at least in part, what might have led him to that view. I will argue that Davidson's view of causal deviance is influenced by a certain conception of the causal structure of action – that is, the structure of the causal relations between actions and their effects, and between actions and the psychological events or states that cause them – and that this conception follows from one natural interpretation of his theory of agency. I will argue, further, that this conception is mistaken. My ultimate conclusion will be that when we correct this conception of the causal structure of action we end up with a conception of the causal structure on which internal causal deviance is a problem that *can* be solved.

1.2.2.1 Davidson's theory of action descriptions

Because of the way in which Davidson approaches the problem of agency, he encounters what I will call the *problem of how many actions*. As a result, a considerable part of his treatment of agency involves developing a solution to this problem. I will first describe how Davidson approaches the problem of agency, will say what the problem of how many actions is and why its solution matters to Davidson's account of agency, and will then reconstruct the solution he proposes. The solution does not go all the way in giving a causal analysis of agency, but it imposes strict, clear limits on what remains to do, and a solution to the residual problem – which I will discuss in the next section – is provided by Davidson's theory of action as I earlier reconstructed it.

As we have seen, Davidson construes the problem of agency as the problem of explaining the agent relation. As is usual for Davidson, the road to ontology passes through ordinary language,³² and Davidson's approach to determining the nature of the agent relation is to consider *action descriptions*, ordinary descriptions of things an agent does - what we ordinarily say to assert the obtaining of the agent relation. We can think of action descriptions as answers to the question, "what did the agent do?"

When we turn our attention to action descriptions, however, we encounter the problem of how many actions: when a person acts on a single occasion, there is in general no unique answer to the question, "what did she do?" Instead, there seem to be many, maybe even endless descriptions. For instance, on a single occasion of acting all of the following seem to be legitimate answers to the question, "what did the man do?": "he flipped the switch," "he turned on the light," "he illuminated the room," and "he unknowingly alerted a prowler."³³ To give a second example, all of the following seem to be legitimate answers to the question, "what did the queen do?": "she moved her hand," "she killed the king," "she emptied a vial into the king's ear," "she killed the king by pouring poison in his ear," "she poured poison in the king's ear thus causing his

death," "she killed the king by moving her hand," "she moved her hand thus causing the death of the king," "she moved her hand thus causing a vial of poison to empty into the king's ear thus causing the king to die."³⁴ The examples make clear, I think, that when a person acts there will typically be many different descriptions of things she does.

The existence of multiple descriptions of what a person does on a single occasion of acting leads to the problem of how many actions. On the one hand, there seems to be an inclination to say that when a person acts she normally performs a *single* action. To cite just one of many possible examples, in a discussion concerning the individuation of actions performed by different agents, Hart and Honoré say that "A's *action* ends where B's *action* begins;...the latter is...a 'new *action*'."³⁵ But, on the other hand, it appears that different action descriptions that apply on a single occasion of acting may attribute different, incompatible properties to the action. For instance, in the first example it is reasonable to suppose that flipping the switch, turning on the light and illuminating the room are all intentional, and alerting the prowler is not. There are also cases in which different action descriptions appear to cover "a smaller or larger stretch of events, those excluded by the narrower description being then called 'consequences' or 'results' or 'effects' or the like of his act."³⁶ This seems to be true of the descriptions, "the queen moved her hand" and "the queen killed the king." On the second view, we can explain the difference between the two descriptions by saying that the action that "the queen moved her hand" describes encompasses a narrower stretch of events than the action that "the queen killed the king" describes, and that some of the later parts of the stretch that "the queen killed the king" describes are effects of the stretch that "the queen moved her hand" describes. How to individuate events is not obvious, but Davidson's assertion that "events that cover different stretches [of events] cannot be identical"³⁷ seems plausible. So though we seem inclined to say that when a person acts she performs a single action, there are also considerations that suggest that when a person acts she performs multiple, numerically distinct actions.

Davidson rejects the conclusion that different descriptions that apply on a single occasion of acting are of different actions, and holds instead that the "welter of related descriptions corresponds to a single descriptum."³⁸ The *single descriptum theory* of action descriptions is a special case of Davidson's view of events, on which events are concrete particulars – chunks of space-time – with multiple descriptions.³⁹ The main motivation for Davidson's extensionalist view of events is presumably his commitment to the extensionalist program – to doing away, in one's philosophical theories, with abstractions, such as properties and propositions, in favor of entities that exist in space and time. But the single-descriptum theory has benefits of its own, and provides further motivation for an extensionalist view of events. For one thing, if the welter of descriptions can be explained in terms of it, they can be explained in a way that preserves the intuition that the agent performs a single action when she acts. Furthermore, the single-descriptum theory not only simplifies the problem of explaining the relation between a person and an event that makes the person the agent of the event, but also provides an elegant and persuasive account of action descriptions, as I will now try to show by developing the theory.⁴⁰

The first problem we encounter in developing the single descriptum theory is that even if we assume that all descriptions of what the agent does when she acts on a single occasion describe a single event, it does not obviously follow that there is a single kind of relation that obtains between the agent and the event, since there are three different locutions we can use to describe what the agent did. Differences in surface form do not, of course, necessarily indicate differences in logical form, but they are *prima facie* evidence for a difference - and so for the existence of three distinct notions or senses of being intentional, one corresponding to each locution. My strategy, suggested by results of Davidson in "Agency," will be to argue that we can express two of the locutions in terms of the third, and so the relations to which the first two locutions refer can be expressed in terms of the relation to which the third locution refers.

Suppose Smith performed the action of opening a certain door. One locution we can use to describe what she did is the *simple locution*: Smith opened the door. The second locution is the *cause-locution*: Smith *caused* the door to open. The third locution is the *by-locution*: Smith opened the door *by* (we suppose) moving her hand. What distinguishes the by-locution from the other two is that it relates a description of an action not just to the agent but to another description as well, often using "by," though other prepositions, such as "through," can also be used. What distinguishes the cause-locution from the simple locution is the use of the specific verb "cause." All three locutions can describe things that happen to a person rather than her actions, but we will consider only cases in which they are used to describe actions. I will now reconstruct Davidson's proposal for expressing the cause-locution and the by-locution in terms of the simple locution.

In action descriptions we describe an action in terms of an event. A lesson of "Agency" is that in many cases the event in terms of which we describe an action is not the action itself but another event, an *effect* of the action.⁴¹ We will thus distinguish the event an action description *describes* - the action itself - from the event the action description *invokes* - the event in terms of which the action is described. For instance, in the case of the action description, "Smith opened the door," Smith's action - her *causing* the door to open - is expressed in terms of an effect of the action - the door's opening. So "Smith opened the door" invokes the door's opening, an event that is an effect of the event the description describes, Smith's causing the door to open, which is the action itself. We use the distinction between the events described and invoked by an action description to define the notions of *primitive* and *non-primitive descriptions of actions*.⁴² If the event invoked in the action description is the action itself, then the description is primitive, but if the event invoked is an effect of the action, then the description is non-primitive. As I have mentioned, Davidson's view is that every action is an intentional movement of an agent's body, and that all other events attributable to an agent are so because they are *effects* of intentional

movements of her body. As a result, only descriptions of actions in terms of bodily movements can be primitive.

Davidson expresses the cause-locution in terms of the simple locution by taking "*A* causes ___" to be elliptical for "*A* does something, which causes ___,"⁴³ thus analyzing the cause-locution in a way consistent with event causation. So to say that the man causes the door to open is to say that the man does something, and what he does causes the door to open. The cause-locution is useful when we know that an agent's action had a certain effect but do not know what the action was, when we want to focus attention on a certain effect of an action and are unconcerned with the action, or when we have some other reason for wanting to talk about an effect of an action while leaving the action itself indefinite. Thus whenever we use the cause-locution we provide a non-primitive description of an action - we describe the action by invoking an effect of the action. For instance, in the description "the man causes the door to open" we describe what the man does that causes the door to open by invoking an effect of what he does, the door's opening. This analysis of the cause-locution enables us to reduce whatever relation between an agent and an event that the cause-locution expresses to whatever relation between an agent and an event that the simple locution expresses, plus the causal relation. Since Davidson's project, as we are reconstructing it, is (at this point) to give a causal analysis of agency, this result marks progress: if we can give a causal account of the relation expressed by the simple locution, we have a causal account of the relation expressed by the cause-locution ready at hand.

Next we consider Davidson's analysis of the by-locution. We note, first, that we can distinguish a causal use of "by" from a non-causal use.⁴⁴ When the "by" is causal, we can rewrite the by-locution using "cause" instead of "by." For instance, we can rewrite "Smith opens the door by moving her hand" as "Smith moves her hand, which causes the door to open." But when the "by" is non-causal, we cannot rewrite the by-locution in this way. For instance, we cannot rewrite "Smith checkmates her opponent by moving her pawn" as "Smith moves her pawn, which causes

her opponent to be checkmated." We will follow Davidson in restricting our attention to the causal use of "by."

Davidson analyzes the (causal) by-locution in terms of the simple locution as follows.⁴⁵ Suppose that Smith opens the door by moving her hand. We will say that she does so if and only if (i) Smith opens the door, (ii) Smith moves her hand, (iii) Smith's opening the door and Smith's moving her hand are the same event, and (iv) the event invoked in the action description "Smith's moving her hand" causes the event invoked in the action description "Smith's opening the door" - in this case, Smith's hand movement causes the door's opening. More generally, we say that A a's by b'ing if and only if (i) A a's, (ii) A b's, (iii) "A a's" and "A b's" describe the same event, and (iv) either the event invoked in "A b's" causes the event invoked in "A a's," or (in the degenerate case) the two events invoked are identical. As in the case of Davidson's analysis of the cause-locution, his analysis of the by-locution enables us to reduce whatever relation between an agent and an event that the by-locution expresses to whatever relation the simple locution expresses, plus other relations consistent with a causal analysis of agency. Thus we can now restrict our attention to the simple locution.

The next problem we face in developing the single descriptum theory arises from the fact that when a person acts there are many descriptions that can be made using the simple locution alone. For instance, it may be true that on a single occasion a man did all of the following: he moved his arm, flicked a switch, turned on the light, and illuminated the room. As we have seen, on Davidson's analysis all of the descriptions describe a single event. But this neither explains their differences nor provides a detailed account of their interrelations.

We can explain the differences among the descriptions in terms of the events invoked. The description "Smith alerts the prowler" invokes the event, the prowler's being alerted; "Smith illuminates the room" invokes the room's being illuminated; "Smith turns on the light" invokes the light's being turned on. All these are non-primitive descriptions of the action, and the events

they invoke are effects of the action. We will assume, however, that in the present example "Smith moves her arm" is a primitive description of the action, and so the event this description invokes is the action itself.

The fact that simple locutions can provide *non*-primitive descriptions of actions shows that there is an alternative to using the cause-locution to assign responsibility to an agent for an event that is an effect of an action she has performed. We can use the cause-locution to specify the effect directly, or we can instead use the simple locution to make a non-primitive description of an action, and specify the effect indirectly. When we specify an effect indirectly using a simple locution, we do not indicate that it is an effect of an action syntactically as we do when we use the cause-locution, but semantically or pragmatically: we use an action description that includes a verb phrase that means the same in the relevant context as a verb phrase that begins with "cause." For instance, we can alternatively express "Smith alerts the prowler" as "Smith causes the prowler to be alerted," "Smith illuminates the room" as "Smith causes the room to be illuminated," and so on.

But can we, in the context of the present example, alternatively express "Smith moves her arm" as "Smith causes her arm to move"? We have assumed that the locution "A causes ___" is short for "A does something, which causes ___." So the question is this: in the context of the example, can we alternatively express "Smith moves her arm" as "Smith does something, which causes her arm to move"? To answer the question, it is important to see that action descriptions that invoke bodily movements of the agent are ambiguous. From a description itself, we cannot tell whether it invokes an action or an effect of an action – whether it is primitive or non-primitive. In some contexts, "Smith moves her arm" is true if taken to invoke an effect of an action - for instance, if Smith moves her (paralyzed) arm by using her other arm. In those contexts, in the sense in which "Smith moves her arm" is true, we can rewrite it as "Smith does something, which causes her arm to move," and so as "Smith causes her arm to move." But in the

present example, we have assumed that “Smith moves her arm” is a primitive description, and invokes the action itself, not an effect of the action. In the context of the example, then, “Smith moves her arm” is false if taken to mean “Smith does something, which causes her arm to move.” So in that context, if one uses “Smith moves her arm” in the sense in which it is true, one cannot alternatively express it as “Smith causes her arm to move.”

To explain the interrelations among the multiple descriptions, we start by noting that we can use the by-locution to order the descriptions: the man alerts the prowler *by* illuminating the room, he illuminates the room *by* turning on the light, he turns on the light *by* flicking the switch, and he flicks the switch *by* moving his hand. We will call a set of descriptions ordered using the by-locution a *by-ordering*. In this particular case the by-ordering is a total ordering since the by-relation is reflexive, antisymmetric, transitive, and relates every description to every other description. However, it is clear that the by-relation does not provide a total ordering for every set of descriptions. For instance, there may be *two* prowlers the man alerts by illuminating the room. Then he alerts prowler 1 by illuminating the room and he alerts prowler 2 by illuminating the room, but he neither alerts prowler 1 by alerting prowler 2 nor alerts prowler 2 by alerting prowler 1. (Recall that we are considering only the causal use of “by.”) But although the by-ordering in this case is not a *total* ordering, it is a *tree* ordering. There is a root element - there is one thing the agent does by which he does everything else. And the ordering converges towards the root - in every case in which there are two different things by which the agent does a third thing, she does exactly one of the first two things by doing the other. For instance, the man alerts prowler 1 by illuminating the room and also by moving his hand, he illuminates the room by moving his hand, but he does not move his hand by illuminating the room.

The next step in ordering the multiple simple descriptions is to apply clause (iv) of our analysis of the by-locution: if A a's by b'ing, then the event invoked in the description "A b's" causes the event invoked in the description "A a's." Accordingly, since the man alerts the prowler

by illuminating the room, the room's being illuminated causes the prowler's being alerted. Similarly, since the man illuminates the room by turning on the light, turns on the light by flicking the switch, and flicks the switch by moving his hand, we can infer that the light's being turned on causes the room to be illuminated, the switch's being flicked causes the light's being turned on, and the movement of his hand causes the switch's being flicked. Thus this procedure enables us to replace the by-relation with the causal relation, and to replace by-trees, whose nodes are *descriptions* of actions, with causal trees, whose nodes are *events*.

This is a significant result. It shows, first, that corresponding to the many descriptions made using the simple locution are many events. However, though all of these events are in some sense attributable to the agent, the result also suggests we can distinguish two reasons why an event might be attributable to an agent: because it is *described* by a simple locution, or because it is *invoked* by one. In the first case we will say that the event is *directly* attributable to the agent, and in the second that it is *indirectly* attributable to the agent. (Since primitive descriptions invoke actions we will stipulate that in their case the event invoked is directly attributable to the agent.) Since every simple locution used to describe what an agent does when she acts on a single occasion describes the same event, the action, only the action is *directly* attributable to the agent. And since simple locutions can invoke many different events, many different events can be *indirectly* attributable to the agent. However, except for simple locutions that are primitive descriptions and invoke an action, all other simple locutions are *non-primitive* descriptions that invoke *effects* of the action. So every event that is indirectly attributable to an agent is an effect of an action. This means we can express an indirect attribution made in a non-primitive description of an action in terms of a direct attribution, plus the notion of cause: in such cases, an event is indirectly attributable to an agent if and only if it is an event that is an effect of an event that is directly attributable to the agent. In other words, except for the action itself, every event attributable to an agent is so because it is an effect of an action. As Davidson puts it, "[w]e may

indeed extend responsibility or liability for an action to responsibility or liability for its consequences, but this we do, not by saddling the agent with a new action, but by pointing out that his original action had those results."⁴⁶

We can now account for the fact that different descriptions of what a person does on a single occasion can describe events that appear to be themselves made up of stretches of events of different lengths, and so that appear to be distinct from one another. This seemed to be the case for the two descriptions, "the queen moved her hand" and "the queen killed the king." On Davidson's theory, the difference between the two descriptions is not that they describe different actions, but that they invoke different events - what "the queen moved her hand" invokes is the action itself, while what "the queen killed the king" invokes is an effect of the action. More generally, what explains the fact that different descriptions of what a person did on a single occasion can describe events that appear to be made up of smaller or larger stretches of events is that different descriptions of the same action can invoke events that are more proximal or more distal effects of the action, as well as the action itself. When the event invoked is a more proximal effect of an action or the action itself, the action appears to be made up of a smaller stretch of events, and when the event invoked is a more distal effect of an action, the action appears to be made up of a larger stretch of events. So we can explain the difference in the size of the stretches of events that different action descriptions seem to describe without assuming that they describe *actions* that encompass smaller or larger stretches of events, and that are themselves distinct events.

We can, in a similar way, account for the fact that different descriptions of what a person does on a single occasion can describe events that appear to be different with respect to whether they are intentional. Consider our supposition that illuminating the room was intentional while alerting the prowler was not. On Davidson's theory of action descriptions, the difference between the two descriptions is not that they describe two different events, one intentional and the other

not: it is not events themselves that are strictly speaking intentional or unintentional, but events relative to some description or other. What is required for an event to be an action is that it be intentional under *some* description, not under every description.⁴⁷ In the case of the description, "Smith illuminated the room," we describe the action as an event that has the property of causing the room to be illuminated, and the action is intentional under this description since the agent intentionally brought about an event with that property. But in the case of the description, "Smith alerted the prowler," we describe the action as an event that has the property of causing the prowler to be alerted, and the action is not intentional under this description since the agent did not intentionally bring about an event with that property. Thus we explain the fact that under some descriptions what a person does is intentional while under other descriptions what she does is not intentional without assuming that the descriptions describe different events, some of which are intentional and some of which are not.

1.2.2.2 A simplification and a solution of the problem of agency

I think we can now appreciate Davidson's claim, at the end of "Agency," that his results lead to "a vast simplification of the problem of agency." It will be recalled that the problem of agency, as Davidson construes it, is to explain the agent relation, the relation between an agent and an event that makes the event an action. The problem seemed very complicated, since when an agent acts, there are normally many different descriptions of things she does. If the different descriptions described different events and expressed different kinds of relations, the relations between the agent and many different events, as well as several different kinds of relations, would need to be explained. But following Davidson, we have explained all the descriptions as describing a single event, *the* action, and all other events attributable to the person as effects of the action.⁴⁸ Further, we have done so in terms of a single instance of a single kind of relation between agents and their actions. Given these results, to give a causal analysis of agency one need

only explain this relation. It is the explanation of all action descriptions in terms of this single relation that, as I understand it, constitutes the “vast simplification.”

One of our goals in reconstructing Davidson’s theory of agency is to see how he can explain the relation between agents and their actions in causal terms. That is, we want to see how he can provide an analysis of sentences of the form, “e is a’s action,” in terms of the causation of the action by an event that involves the agent’s instantiation of some property.⁴⁹ Though we have not yet achieved this goal, we have characterized the relation, in our definitions of action (A) and of agency (AG) taken together, as the relation of being intentional. The simplification of the problem of agency also shows that we attribute to agents not just their actions, but some of the effects of their actions.⁵⁰ We can explain the effects for which we hold agents responsible as *being intentional as an effect*,⁵¹ a notion derived from our notion of being intentional:

(IE) An event e is intentional as an effect under a description q for a person p if and only if there is a reason r to make it the case that q such that:

- (a) p has r,
- (b) r causes an action a of p,
- (c) a causes e.⁵²

The following example will illustrate the definition. Suppose that the event, the door’s being opened, is intentional as an effect for Smith under the description, “the door’s being opened.” By (IE) this is so since (we suppose) there is a reason to make it the case that the door is opened such that (a) Smith has the reason, (b) Smith’s having the reason causes an action of Smith’s, her moving her hand, and (c) Smith’s moving her hand causes the door’s being opened.

We can use the definitions of being intentional and of being intentional as an effect to define a general notion of *being responsible*:⁵³

(P) A person p is responsible for an event e iff:

(a) e is intentional under a description for p , or

(b) e is intentional as an effect under a description for p .

When the relation of being responsible holds between a person and an event, we may hold the person liable for the event, or blame or praise her for making it occur. Thus we may praise or blame a person for some of the effects of her actions, and for all of her actions.

As I mentioned, the simplification does not itself provide a causal analysis of the relation between the agent and her actions, and Davidson does not give such an account in “Agency.” However, we can combine the simplification with definitions we have already stated to give such an account. By the definition of agency (AG), a person is the agent of an event if and only if the event is her action. By the definition of action (A), an event is her action if and only if it is intentional under some description for her. And by the definitions of being intentional under a description (I) and having a reason (R), an event is intentional under a description q for her if and only if her having a desire that q , and her having a belief that she has a way to make it the case that q , cause the event to occur, and the event is a movement of her body. Thus, following Davidson, we have explained the relation between an agent and her actions as a causal relation between an agent’s having a reason to move her body and a movement, where having the reason causes the movement. This result is consistent with a causal analysis of action: an action is a movement of an agent’s body if and only if it is caused by her having a reason to make the movement. So we have achieved our goal of explaining actions as events that have a certain kind of cause, and what now remains to achieve our ultimate goal of giving a physicalistic reduction of action is to give physicalistic reductions of the causes of actions, beliefs and desires. I will sketch physicalistic reductions of belief and desire in sections 2.4.1 and 2.4.3 of the next chapter, and

will develop a more detailed physicalistic reduction of a component of belief, indication, in chapter 3.

1.2.3 Davidson's response to causal deviance

I will now consider how the problem of causal deviance bears on Davidson's simplification of the problem of agency and the causal analyses of agency and action, and how Davidson responded in his theory of action to the problem of causal deviance. Though elsewhere in this chapter I focus on the problem of internal causal deviance, it will be instructive to consider here the problem of external causal deviance as well. Because of the possibility of internal causal deviance, a person's having a reason to make a movement might cause her to make the movement, but the movement might fail to be intentional since her having the reason might cause her to make the movement in a deviant way. For instance, in the example of internal causal deviance I gave earlier, the mountain climber's having a reason to open her hand to let go of the rope supporting her partner causes her to open her hand, and so by the definition of being intentional (I), opening her hand is intentional, and is her action. But we are clearly much less inclined to ascribe to the mountain climber responsibility for opening her hand, and to say that opening her hand is her action, in a case like this, in which her reason causes her movement in a deviant way, than if her reason causes her movement in a normal way. So we cannot reduce the relation between a person and her actions to a causal relation between her having certain psychological states and making certain movements. Similarly, because of the possibility of external causal deviance, a person's having a reason to make it the case that a fact obtains might cause an action (in a normal way), and the action might cause an event in virtue of which the fact obtains, but the event might still fail to be intentional as an effect because the action might cause the event in a deviant way. So we cannot reduce the relation between a person and the intended

effects of her actions to a causal relation between her actions and their effects, plus the relation between the agent and her actions. As Davidson puts it,

not just any causal connection between rationalizing attitudes and a wanted effect suffices to guarantee that producing the wanted effect was intentional. The causal chain must follow the right sort of route.⁵⁴

So we must reject the solution implicit in Davidson's theory of action to the simplified problem of agency, and we must reject the simplification as well.

To characterize Davidson's response to the problem of internal causal deviance,⁵⁵ I will modify the definition of being intentional (I) as (I') by adding a clause stipulating that the causal chain from the agent's having a reason to make a movement to her making it is normal, and I will make a similar addition to the definition of being intentional as an effect (IE) to get (IE'). (Note that to deal with the problem of external causal deviance, we would need to add another clause to (IE) stipulating that the causal chain from the action to the effect of the action is normal.) To complete a causal analysis of (I') and of (IE'), we must say what the difference between deviant and normal causal chains is in simpler, strictly causal terms. Since Davidson thinks this is impossible, and the notion of causally normal must be primitive, being intentional is only *partly* analyzable in such terms, and there is a residual unanalyzable element. A consequence is that we cannot give sufficient causal conditions for being intentional, and for action. It is not sufficient for a movement to be intentional, and an action, that it be caused by an agent's having a reason to make it. The way in which the movement is caused must be further constrained, but Davidson thinks it is not possible to say how it must be constrained in physicalistic terms. Thus, as a result of the change in the definition of being intentional, we can no longer give a causal analysis of action. And, if we cannot do that, we must abandon our strategy for giving a physicalistic

reduction of action of first giving a causal analysis of action and then expressing the causal analysis in physical terms. So, if we accept the revised definition of being intentional, we cannot follow our strategy for explaining the relation between the agential and the non-agential in a way that leaves the world conceptually as well as ontologically unitary. On the modified definition, conceptually, the world is divided between the agential and the non-agential.

Treating the notion of causally normal in (I') as primitive leaves us with a theory of action that still explains many things about action and agency, such as the relation between agents and their actions, and the difference between actions and passive happenings, even if it no longer provides a causal analysis of those notions. And even if we cannot analyze the notion of causally normal in simpler, physicalistically acceptable terms, we can still explain the notion informally, by means of examples, as we have done. So Davidson's theory of action is attractive even if we treat the notion of causally normal as primitive. On the other hand, treating causally normal as primitive leaves us with two different ways of talking about the world, at the level of actions and agents, and at the physical level; and if we have two different ways of talking about something, it is better on theoretical grounds if we can relate them. Further, an analysis of the notion of causally normal is preferable to an informal account on the basis of examples because, though an informal account may provide a basis for classifying cases that everyone agrees are normal or deviant, it cannot provide a basis for classifying cases about which people's intuitions differ; nor does it explain *why* any case is, or is not, a case of causal deviance. So if we can explain the notion of causally normal in physicalistic terms, we will not only provide causal analyses of action and agency, and pave the way for physicalistic reductions of those notions, but we will improve Davidson's theory of action. This is the task I will take up in the next section.

1.3 A solution to the problem of internal causal deviance

I will now describe a proposal for dealing with the problem of internal causal deviance, and will suggest how to apply it to explain why some movements are the result of internal causal deviance and the agent is not fully responsible for them, or not responsible at all. Later, in chapter 2, I will formulate, illustrate and defend the proposal in detail. Here, however, my aim is to give a brief, impressionistic idea of how it works and what resources it requires. This will set the stage for considering, in section 1.4, why Davidson might have overlooked the possibility of such a proposal, and thought the problem of internal causal deviance insurmountable.

Central to the proposal is the notion of a *modulated movement*. What characterizes a modulated movement is that it is extended through time to a degree that allows the person making it to receive and respond to feedback about the course it takes as she carries it out.⁵⁶ Modulated movement is a form of trial-and-error. For instance, when a baseball player catches a fly ball, she generally makes a series of attempts to position her body and glove so that she can catch the ball: she positions her body and glove in a first way, refines the first position on the basis of her observation of the trajectory of the ball, refines the second position on the basis of further observation, and so on, until – if all goes well – the ball lands in her glove.

As the example suggests, carrying out a modulated movement is an iterative process that can be divided into three component activities: (1) *information gathering*, (2) *decision making*, and (3) *movement modulation*. (I will consider the component activities in more detail in section 2.2 of the next chapter.) I mentioned earlier that an action, on Davidson's view, is a movement of an agent's body caused by her having a reason to make the movement. For an action to involve modulated movement, the agent must continue to have the reason as she carries out the movement, and the reason must guide her in carrying out the component activities. The key idea underlying the notion of modulated movement is that when an agent acts by carrying out a

modulated movement, she does not just initiate the movement and then wait for it to run its course, but instead keeps acting until she achieves the desired result, and the action is complete.

What is crucial for my proposal for dealing with the problem of internal causal deviance is this. When a person's having a reason for a movement causes her to carry out the movement in a normal way, the movement is properly modulated. But when a person's having a reason for a movement causes her to carry out the movement in a deviant way, the modulation is defective or absent because of her failure to carry out properly at least one of the component activities required for modulated movement. For the mountain climber to let go of the rope in a normal way, she would have to keep doing all of the following as she lets go of the rope: observe how she grips the rope with her hands, make decisions about how to move her hands based on what she observes, and move her hands in accordance with what she decides. Letting go of the rope would be deviant if, for instance, she fails to observe how she grips the rope, cannot make decisions about how to move her hands based on what she observes, or trembles so violently that she cannot adjust her grip on the rope in the way necessary to let go of it in an intelligent, systematic way.

I hope to have said enough in this section to suggest that there is a promising strategy for explaining the difference between movements caused in deviant ways and ones caused in normal ways. If I am right that deviant movements can be distinguished from normal ones by considering whether they are properly modulated, then we can say in general, in physicalistic terms, how deviant internal causal chains are different from normal ones (provided we can explain the relevant psychological states in physicalistic terms), and the problem of internal causal deviance is not a reason to abandon the project of giving a physicalistic reduction of action.

1.4 Diagnosis of Davidson

I want to consider now why Davidson might think the problem of internal causal deviance insoluble. If there is a promising strategy for dealing with the problem, as I have suggested, why is he certain no strategies can be found? In particular, what might have led Davidson to overlook the strategy of distinguishing deviant from normal movements by considering whether they are properly modulated?

The reason, I think, is that he holds a mistaken conception of action, the *ballistic conception*. On this conception, there can be no such thing as modulated movement. Instead, all actions involve *ballistic movements*, movements the agent initiates but cannot subsequently modify or control. (Later, in section 2.4.2 of the next chapter, I will discuss ballistic movement in detail, and will argue that the distinction between normal and deviant movement does not apply to it.) In section 1.4.1 I will describe the ballistic conception, and how it follows from a particular conception of causation, the conception of causation as *wholesale*.⁵⁷ This conception of causation seems quite natural, and if Davidson thinks all causation is wholesale, that could explain why he overlooked the strategy of explaining the difference between deviant and normal movements in terms of their modulation, since, as I will argue, modulated movement is impossible if all causation is wholesale. In section 1.4.2 I will consider an alternative conception of causation, the conception of causation as *piecemeal*.⁵⁸ While movements that are wholesale-caused - *w-caused* - must be ballistic, movements that are piecemeal-caused - *p-caused* - can be modulated. My contention will be that when we consider certain kinds of phenomena, it is natural to think that the causation involved in them is piecemeal, and that action should be assimilated to those phenomena.

1.4.1 *The ballistic conception of action and wholesale causation*

I will now describe the ballistic conception of action, w-causation, and the relation between them. As we saw in section 1.2.2.2, Davidson's simplification of the problem of agency, together with his causal theory of action, leads to the view that when a person acts there is a tree of causes and effects. The sequence begins with an agent's having a reason to make a movement, which causes the next event in the sequence, the movement. The movement, in turn, sets off a tree of events, which may extend indefinitely. So far, this picture is consistent with both the ballistic conception of action and the conception of action as modulated movement. What characterizes the ballistic conception is its inclusion of the thesis that the causation of one event by another event is invariably wholesale: that is, for ordinary or practical purposes, the causation itself is an event that may be regarded as instantaneous, with no structure, and is confined to *setting off* the second event. Once that has happened, the first event ceases to have any effect on the second event. We will say that events like the causation, which have no recognized structure, are *brief*, and will contrast them with events that we think of as processes consisting of many smaller events, which we will say are *prolonged*.⁵⁹

To illustrate the ballistic conception of action and the notion of w-causation, I will first describe a simple example that does not involve action, and will then extend the example to the case of action. A paradigm case of w-causation is one billiard ball's causing the movement of another by colliding with it. Here one process, the movement of the first ball, causes another process, the movement of the second ball. We can think of the causation of the movement of the second ball by the movement of the first ball as the event that consists of the transfer of energy from the first ball to the second ball – their collision – so the causation is a nearly instantaneous event whose occurrence is confined to the time when the balls are in contact, and ends when the second process has been set off. As a result, the movement of the first ball has no further effect on the movement of the second ball after setting it off.

Now let us extend the example to a case of action. Suppose again that one process causes another process, but that the first process is an agent's having a reason to move her hand, and the second process is the movement of her hand. On the ballistic conception of action, the causation of the agent's hand movement by her having a reason to move her hand has the same structure as the causation of the movement of the second billiard ball by the movement of the first billiard ball in the preceding example: the causation is a brief event that is confined to setting off the hand movement. Once the agent's having a reason to move her hand has set off the hand movement, having the reason ceases to have any effect on the movement, and the intelligent role of the agent in the movement is over.

We are now in a position to see why the ballistic conception of action is inconsistent with the view that actions involve modulated movement, and so forecloses explaining deviant movements as movements that are improperly modulated. Proper modulation requires that the event that consists of the causation of the movement by an agent's having a reason to make it include her monitoring and responding to the course of the movement as she carries it out. But on the ballistic conception that is impossible, since the event that consists of the causation is a brief event that is over once the movement starts. It follows that it is impossible for agents to modulate their actions as they carry them out to respond to new incoming information about the actual course of the movement. Though an agent's movements are prolonged, and their effects can extend indefinitely, once an agent has set her body into motion she has no control over her movements or over any of their effects, and there can be no such thing as modulated movement.

In the next section I will describe p-causation, which is consistent with modulated movement. I will contrast it with w-causation, and will argue that we should view the causation involved in typical cases of action as piecemeal.

1.4.2 Wholesale causation and piecemeal causation

As we have seen, when the causation of one process by another process is wholesale, the causation is a brief event and is confined to the first process's setting off the second process. Once that has happened, the first process has no further effect on the second process, so only one part of the second process, its beginning, is caused by a part of the second process. By contrast, when the causation of one process by another process is piecemeal, the causation is prolonged, and is *not* confined to the first process setting off the second process. Instead, even after the second process has been set off, the first process continues to have effects on the second process so that every part of the second process is caused by some part of the first process. As a result, different parts of the first process can cause different parts of the second process at different times. To sharpen the contrast, suppose that a process A consists of the sequence of events a_0, a_1, \dots, a_n and that a process B consists of the sequence of events b_0, b_1, \dots, b_m , where m is less than or equal to n . If A w-causes B, then some part a_i of A causes the first part of B, b_0 , and this sets off a chain reaction, so that b_0 causes b_1 , b_1 causes b_2 , and so on, until b_{m-1} causes b_m . But if A p-causes B, some part a_i of A causes the first part b_0 of B, a part of A that is later than a_i causes the second part b_1 of B, and so on, until some part of A causes b_m . P-causation is consistent with B feeding back to A – which is required for modulated movement – but is not sufficient for it, since for there to be such feedback, it must further be true that some part of B causes a part of A that is later than the first part of A that causes a part of B.

I will now illustrate the distinction between w-causation and p-causation with an example that uses historical events. Consider the sentence

- (1) A bombing campaign caused massive destruction.

If what is meant is that the causation of the second process, the destruction, by the first process, the bombing campaign, is wholesale, then (1) asserts that some particular part of the bombing campaign set off all of the destruction. For instance, the part of the bombing campaign might be the bombing of a munitions factory located in a large, densely populated city, and the destruction might be the burning of the city that began when bombs dropped on the munitions factory started a fire that spread to the rest of the city. But if what is meant is that the causation of the destruction by the bombing is piecemeal, then (1) asserts that different parts of the bombing campaign caused different parts of the destruction at different times. Thus the bombing of the munitions factory might have caused one part of the destruction, the bombing of another munitions factory in another city might have caused another part of the destruction, the bombing of a dam might have caused yet another part of the destruction, and so on, so that every part of the destruction was caused by some part of the bombing campaign. It seems clear that there are many cases in which it is more natural to interpret (1) as asserting p-causation rather than w-causation, and that many sentences that assert that one event causes another event, such as those that refer to historical events, are like (1) in this respect.

Now let us consider the distinction between w-causation and p-causation as it applies to action. Consider the sentence

(2) Smith's having a reason to move her hand caused her to move her hand.

If what is meant is that the causation is wholesale, then (2) asserts that some particular part of Smith's having the reason – say, her having the reason at a particular time – caused the first part of her hand movement, which set off a chain reaction in the relevant parts of her body. So the first part of the movement caused a slightly later part, which caused an even later part, and so on, until the movement was complete. A chain reaction like this would occur if having the reason set

Smith's hand into motion, and the rest of her hand movement were the result of the momentum from her hand first being set into motion. By contrast, if what is meant is that the causation is piecemeal, then (2) asserts that different parts of her having the reason caused different parts of her hand movement at different times, so that progressively later parts of having the reason caused progressively later parts of her hand movement until every part of her hand movement was caused by some part of her having the reason.

When having a reason to make a movement causes the movement in a case of action, is it more plausible that the causation is piecemeal or wholesale? Later, in section 2.4.2 of the next chapter, I will argue at length that all actions involve modulated movement, and that ballistic movements – movements that are w-caused rather than p-caused by the reason for the movement – can be intentional only in the sense that they are intentional effects of actions.⁶⁰ Here I will consider a single example of an action, and will contrast what the action would be like if it were p-caused with what it would be like if it were w-caused. The example is intended to illustrate further the distinction between w-causation and p-causation as it applies to action, and the distinction between modulated and ballistic movement. But it is also intended to show that it is much more plausible in this case that the causation is piecemeal rather than wholesale, and so to make a *prima facie* case that actions involve modulated movement and are p-caused by the reasons to make them, rather than involving ballistic movement and being w-caused by the reasons.

In the example, a person performs an action of moving her hand in order to open a door. For simplicity, we suppose that opening the door does not require turning the doorknob, but is a matter of swinging the door open to the desired position. On the ballistic conception of action, not only is the causation of the hand movement by the person's having a reason wholesale, but so is the causation of the door's opening by the hand movement. Then some part of the hand movement causes the first part of the opening, which sets off a chain reaction that culminates in

the door being opened to the desired position. But it seems clear that this is not how we ordinarily open doors. Opening the door in this way would require that an agent's hand have an effect on the door analogous to the effect that one billiard ball has on another when the first ball collides with the other ball, so once the door starts to open the agent's hand has no further effect on it. For this to be the case, the person would have to punch the door with her fist or slap it with her hand. But this is obviously not how we ordinarily swing doors open, and if a person did open a door in this way, I think that we would view her as having little control over how it opened, and that we would not view opening the door in this way as a typical case of action.

A more plausible scenario is one in which the hand movement p-causes the door to open, so that progressively later parts of the hand movement cause progressively later parts of the door's opening. This would be the case if, say, an agent opens the door by tapping it, or by holding the doorknob and swinging the door open. In such cases, it also seems plausible that the agent's having a reason to move her hand would p-cause her moving her hand, since it seems plausible that as she carries out successive parts of the movement she would monitor the course of the door's opening and would modulate the movement of her hand in accordance with the feedback she receives about the movement of the door. If she did not, then she would neither know when the door was opened to the position she desired, nor would she stop moving her hand when the door reached that position.

I have argued that sentences that assert that one event causes another are ambiguous, since they can be interpreted as asserting either that the causation is wholesale or that it is piecemeal. If one takes events like collisions between billiard balls as paradigms of causation, it is natural to think that causation in general is wholesale. One who thinks this will be led to the ballistic conception of action, a conception that forecloses the strategy I have proposed for distinguishing normal from deviant movements, in terms of whether the movements are properly modulated, since it rules out the possibility that agents can monitor and modify their movements

as they carry them out. Thus if Davidson assumes causation is invariably wholesale, and action invariably ballistic, that could explain why he fails to consider the possibility that actions are modulated movements, and that internally deviant movements can be distinguished from normal ones by considering whether they are properly modulated.⁶¹ I have argued, however, that in some cases, such as many cases of historical causation, sentences that assert that a causal relation holds should be interpreted as asserting that the causation is piecemeal, and that sentences that assert that a causal relation holds between an agent's having a reason and an action should be interpreted this way. P-causation is consistent with the conception of action as modulated movement, and so is consistent with the solution I have sketched to the problem of distinguishing normal from deviant movements. If my proposal or one like it is successful, then the problem of internal causal deviance does not pose an insuperable obstacle to the project of giving a physicalistic reduction of action. In the next chapter, I will consider the proposal in more detail.

Chapter 2

Modulated Movement

2.1 Introduction

My aim in the last chapter was to set up the problem of internal causal deviance in the context of Davidson's project of giving an account of action, to sketch an analysis of modulated movement that provides a solution to the problem, and to explain why Davidson might have thought the problem insurmountable. I suggested that the analysis provides a solution to the problem of internal causal deviance because, on my view, an internal causal chain, a chain from a reason to move in a certain way to moving in that way, is normal if and only if the movement is properly modulated. In this chapter my aim is to develop the analysis of modulated movement in more detail and to defend the analysis against some objections that it faces. I present the analysis in section 2.2, and discuss how it copes with internal causal deviance in section 2.3. In section 2.4 I formulate three objections to the analysis and defend it against them. If the analysis is successful - that is, if it provides a true, general explanation of the difference between normal and deviant internal causal chains in physicalistic terms - then the problem of internal causal deviance poses no threat to causal theories of action, and, contra Davidson, does not provide a reason to think that no systematic physicalistic reduction of action can be given. A subsidiary aim of section 2.4 is to sketch reductive physicalistic theories of belief and desire, and to suggest how human actions might be explained in physicalistic terms. In the next chapter I will develop in more detail a physicalistic theory of a component of belief, indication

2.2 Analysis

Recall, from section 1.3 of the last chapter, that a movement is properly modulated (roughly) if and only if, as the agent carries it out, she receives feedback about it and responds to

the feedback by modifying it (or not) as needed to satisfy the desire that forms a part of her reason to make the movement. As I mentioned, modulated movement is a form of trial-and-error. A person tries to attain a goal by moving in a certain way, and as she moves she observes her progress. Depending on what she observes, she makes any changes to how she moves that she needs to in order to attain her goal. I will illustrate these features of modulated movement with another example.

Imagine that a person is walking over an uneven surface – say, she is hiking across rugged terrain from one point to another point. To do so, the hiker does not plan every detail of every step before she sets off. Such planning would require determining in advance, for every step, precisely how to position her feet, and hold her arms and body, in order to maintain her balance and continue towards her destination. She would have to anticipate every eventuality that will befall her – every slip and stumble, every rock and fallen tree blocking her path, and so on. In other words, she would need complete information about her traversal before she started it, and would need to predict everything she would encounter completely and accurately. It is clearly implausible that she could plan her traversal down to this level of detail; for one thing, the computations she would have to carry out are intractable. Instead, she determines to a large extent how to move to traverse the terrain as she traverses it, in “real-time,” on the basis of local information she obtains by an ongoing process of observation. She feels herself start to slip on a rock, makes the decision to jump off it, and does so. When she lands, she feels herself stumble, and chooses to extend her arms to help keep her balance. When she does so, she does not determine in advance how to extend them, but keeps moving them until she is stable. When she encounters terrain that is especially rough, how she needs to move to continue her traversal will be especially unclear. She will have to step very gingerly, observe what happens very closely, and pause frequently to figure out what to do next. She may have to try a couple of different ways of setting down a foot before she puts her weight on it. If something she tries to do moves her

further from her goal rather than closer to it, she will try something different. Because she relies on trial-and-error, she need not predict everything that will occur during her traversal, but can act under considerable uncertainty.⁶²

It is worth emphasizing that I am not claiming there is no advance planning whatsoever when one carries out a modulated movement. The agent may, before she starts her traversal, plot a course over the terrain, and she may, before she takes a step, think about where to place her foot and how to position it. But the plans are partial.⁶³ They must be, since, as I have mentioned, determining in advance, down to the last detail, how to move her feet, legs, arms and body would impose an impossibly heavy computational burden on her. Further, her plans may be based on incorrect information, or incorrectly predicted results – she may think a rock is less slippery than it is, and she may fail to predict that she will slip when she steps on it. Even when the agent plans in advance, she fills in the plans and corrects them as she acts, on the basis of relatively local information, information she receives about the course of her movement as she carries it out.

As I suggested earlier (in section 1.3 of the last chapter), and as the example I just gave illustrates, we can divide modulated movement into three component activities: (1) information gathering, (2) decision making, and (3) movement modulation. I will describe each in turn.

(1) Information gathering is a matter of forming beliefs about the relevant features of the circumstances in which the agent carries out her modulated movement. The agent uses the information she gathers to make decisions about how to modulate her movement in order to satisfy the desire that, in part, caused the movement. Typically, as an agent carries out a modulated movement, her circumstances change, and she must update her beliefs to reflect those changes. One very important feature of her circumstances is the movement itself, since how she should modulate her movement depends crucially on how she is moving, and in many cases an agent must move in certain ways to enable her to move in certain other ways.⁶⁴ Many of the agent's beliefs about her circumstances are perceptual beliefs, beliefs based directly on what she

observes, but her beliefs may also include beliefs formed by drawing inferences from perceptual beliefs to other beliefs about her circumstances. For example, as the hiker traverses the terrain, she will feel or see such things as that she has raised one foot a certain distance off the ground, that her toes have bumped against a rock, or that one of her feet has slipped. On the basis of these perceptual beliefs, she may infer, for instance, that if one foot slipped on a rock when she stepped on it, it is likely the other will.

(2) Decision making, or deliberation, is a matter of discovering, evaluating and choosing options for modifying the ongoing movement in order to satisfy the desire that (in part) causes the movement. What options an agent considers and chooses depend on her beliefs about the course of the movement. For instance, if as the hiker sets a foot down she feels it slip, she may decide that shifting her weight back to her other foot is the best way to satisfy her desire to traverse the terrain. We can assimilate decision making to a widespread phenomenon of theory formation and evaluation. I will develop this idea shortly, when I compare modulated movement to the development of scientific theories.

(3) Movement modulation is a matter of carrying out the decisions the agent makes about how or whether to modify the ongoing movement in order to achieve her goal in carrying it out. In many cases what the agent decides will best contribute to achieving her goal will be to keep moving in the same way rather than modify the movement. For instance, if the hiker is lifting a foot to take a step in order to traverse the terrain, what she may decide will best achieve her goal of traversing the terrain may be to keep lifting her foot to complete the step she is taking. The results of modifying, or not modifying, her movement are to a large extent what the agent forms beliefs about in the data gathering activity of modulated movement.

It is instructive to compare modulated movement to another form of trial-and-error, the development of scientific theories. On the traditional view, developing a scientific theory is a matter of carrying out an iterative procedure of acquiring data about relevant parts of the world,

forming and evaluating hypotheses to explain the data, and testing the hypotheses, where the results of testing the hypotheses provide further data. We can view the information gathering component of modulated movement as an analog of data acquisition for a scientific theory, and the decision making component of modulated movement as an analog of hypothesis formation and evaluation. But modulated movement, unlike the development of scientific theories, is a practical rather than a theoretical business. That is, the ultimate aim of modulated movement is to have an effect on the world, to make something happen – such as to traverse a piece of ground – and not to develop a theory to explain or predict some feature or aspect of the world. This might suggest that we should think of the movement modulation component of modulated movement as a matter of carrying out rather than testing the decisions the agent makes about how to move. But this conclusion is wrong. There are times when an agent may decide that the best way to achieve her ultimate goal of having a certain effect on the world is to make a modification to a movement purely to see what happens when she does. By modifying it in that way, she draws information she can use in making decisions about how to move. For instance, in walking over the terrain, the hiker may be unable to tell by looking whether a surface is slippery, and may set a foot down on it lightly in order to tell. Or she may be unable to tell just by looking which of several possible toe holds will be the most secure, and may try them all before she picks one and continues on her way. Though cases like this, in which an agent modifies a movement purely to acquire information, are relatively uncommon, movement modulation serves the function of drawing information as well as making things happen in more common cases as well. Even in walking over relatively flat ground, each step the hiker takes not only propels her forward, but also tests the decisions she has made that she can keep moving forward by walking more or less normally. So modulated movement resembles scientific theory development in its reliance on testing the conclusions reached, but in modulated movement this testing is subordinate to the goal of having effects on the world.

The picture of action I have presented forms the basis for my analysis of internal causal deviance.⁶⁵ My contention is that when an agent's reason causes a movement in a normal way, the movement is properly modulated; that is, as an agent moves she engages in a process of constantly acquiring new, relatively local information about the course of the movement, determining whether and how to modify the movement on the basis of the new information, and modifying or not modifying the movement on the basis of what she determines. The process is iterative so that the agent keeps carrying out the process until the desire that originally motivated the movement is satisfied. (This is what makes modulated movement a form of trial and error.)

The definition of modulated movement I will propose forms a part of the definition of being intentional (I') (stated in section 1.2.3 of the last chapter). Recall that I formulated (I') to characterize Davidson's response in his theory of action to the problem of internal causal deviance, and that (I') is the same as the initial definition of being intentional (I) (stated in section 1.2.1 of the last chapter), with the addition of a clause that explicitly excludes deviant movements:

(I')(d) the causal chain from the agent's having the reason *r* to the movement *e* is normal.

I take the relation between a movement's being caused by a reason in a normal way and its being properly modulated to be this: a movement is caused in a normal way if and only if it is properly modulated. (Later (in section 2.4.2) I will argue that ballistic movements, movements that cannot be modulated, cannot be actions.) So my analysis of modulated movement, which will provide necessary and sufficient conditions for a movement to be properly modulated, will also provide necessary and sufficient conditions for clause (d). But not every movement that is not properly modulated, and so not normal, is deviant. A mountain climber might form a momentary reason to

let go of a rope, and the reason might cause her to panic. But her panic might cause her to move so that she gets entangled in the rope and secures it. Then her reason does not cause her to let go of the rope in a deviant way since it does not cause her to let go of the rope at all. Therefore I take the causal chain from a reason to a movement to be deviant if and only if condition (d) is not satisfied but conditions (a) through (c) of (I') are (so an agent's reason for a movement of her body causes the movement).

Because I will state my analysis of modulated movement in terms of certain kinds of beliefs, desires, and reasons – more precisely, beliefs, desires and reasons that have certain kinds of contents – it will be useful to have a term to distinguish the reason *r* in the definition (I'), the reason that is primarily responsible for causing the movement that is the action, from the others. I will say that *r* is the *guiding reason*,⁶⁶ and the belief and the desire that make it up the *guiding belief* and the *guiding desire*.

In the last chapter, in section 1.4, I distinguished two kinds of events, brief and prolonged, and I also distinguished two kinds of causation, wholesale and piecemeal. I argued that these two distinctions relate to modulated movement in the following way. When we say, in the case of a modulated movement, that a person's having a reason to make the movement causes the movement, we mean that the person's having the reason and the movement are both prolonged events, and that the causation of the movement by the reason is piecemeal. Accordingly, we state the first part of our definition of a normal causal chain from a reason to a movement as follows:

(I')(d) the causal chain from a reason to a movement is normal if and only if:

- (i) *e* is a prolonged event,
- (ii) the agent's having *r* is a prolonged event.

Where the context does not make clear whether a movement is one of the smaller events that make up the prolonged event *e*, or is *e* itself, I will call the smaller events *submovements*, and *e* the *total movement*. Note that even if *e* is brief it can still satisfy conditions (a), (b) and (c) of (I'). Note too that the causation of *e* by *S*'s having *r* must be wholesale if either *e* or *p*'s having *r* is brief, but can be either wholesale or piecemeal if both *e* and *p*'s having *r* are prolonged.

The next step in stating the definition is to specify additional psychological states besides the guiding reason that an agent must have to carry out a modulated movement. We start by recalling that in the information gathering activity of modulated movement the agent forms beliefs about relevant features of her circumstances, including the movement itself, and constantly updates those beliefs as she carries out the movement. The following clause requires that the agent have those beliefs:

- (d)(iii) there is a prolonged event that is the agent's having beliefs about her circumstances.⁶⁷

I will call these beliefs *circumstantial beliefs*.

Next we consider the psychological states the agent forms when she makes decisions about how to modulate the total movement in order to satisfy the guiding desire. There are two relevant states, which I will consider in turn. I suggested earlier that one thing that decision making involves is the identification and evaluation of ways of modulating the total movement in order to satisfy the guiding desire. Here the relevant psychological states are beliefs that the agent can do one thing by doing another – in particular, that she can satisfy the guiding desire by moving in a certain way. How she needs to move to satisfy her guiding desire obviously depends on her circumstances – on what stage she has reached in carrying out the total movement. So the beliefs the agent must have are beliefs whose content has the following form: in circumstances *c*,

the agent can satisfy the guiding desire by moving in a way m . For instance, the hiker will believe that she can traverse the piece of ground by lifting her right foot over a rock at a certain stage of her traversal only if she also believes, say, that her left foot is firmly planted. So we state the next clause of the definition as:

- (d)(iv) there is a prolonged event of the agent's having beliefs about ways to move in order to satisfy her guiding desire given her circumstances.

I will call these beliefs *by-beliefs*.

In the decision making component of modulated movement, I have suggested, the agent identifies and evaluates various options available to her to continue carrying out the total movement and contribute to satisfying the guiding desire. I also suggested that we can think of this identification and evaluation of options as forming and evaluating hypotheses about how to modulate her total movement, about what submovements to make, so that she can satisfy her guiding desire. There are two ways an agent can evaluate these hypotheses: by actually modifying the total movement and seeing what happens, or by just thinking about doing so, and considering the possible consequences. We can expect there to be a range of cases. In some, the cost of making a mistake is low, or what modification the agent should make is obvious, and she will evaluate very few hypotheses before acting on one. In other cases, the cost of making a mistake is high, or what modification she should make is unclear, and she will need to give more thought to how to modulate the movement, and evaluate more hypotheses or evaluate them more carefully before she acts on one. In either case, in order to modify the movement, she will have to pass from some of her by-beliefs to reasons to move in the ways those by-beliefs describe. Suppose that the hiker has settled upon the option of lifting her right foot over a rock as a promising way at a particular stage of traversing the piece of ground to achieve her goal of completing the traversal

– that is, she has decided that her belief that she can satisfy her guiding desire by lifting her right foot over a rock has the most to be said for it (or, at least, no less than any other belief). Then she will form a desire to lift her foot over the rock. Now, for her to have regarded lifting her foot over the rock as a promising option, she must already have formed the belief that she has a way to lift her foot over the rock. So, in forming the desire, she forms a reason to step over the rock. We will require the existence of reasons about how to modify the total movement - the product of the decision making activity of modulated movement – in the next clause of our definition:

(d)(v) there is a prolonged event that is the agent's having reasons about how to move.

I will call these reasons *subreasons*

We now have clauses in our definition that require the existence of all the psychological states needed for modulated movement. The final step in stating the definition is to say what the causal relations among the events specified in the definition must be for a movement to be modulated.

We start with the relation between the circumstances in which the agent carries out the movement, which include the movement itself, and the agent's having beliefs about the circumstances. It is possible that some or all of those beliefs are not caused by her circumstances – perhaps she is hallucinating – but it is clear that when things are working as they should, her circumstances cause her to form her beliefs about them. Since the agent keeps updating her beliefs about her circumstances as she carries out the movement, different temporal parts of her circumstances cause different temporal parts of her having beliefs about her circumstances, so the causation is piecemeal. Consequently, we specify the relation between the circumstances and her having circumstantial beliefs as follows:

(d)(vi)(a) the circumstances in which the agent carries out the total movement (which include the total movement) p-cause the agent's having circumstantial beliefs.

I characterized decision making as involving two processes, the formation of by-beliefs, beliefs about ways to move, and the formation of subreasons, reasons to move in certain ways. How the agent should move at a particular stage of the total movement depends, I have noted, on what her circumstances are like at that stage, including how she has most recently modulated the total movement. So every belief about how to move should be caused by some belief about her circumstances. Also, since she will need to revise her by-beliefs constantly to reflect changes in her circumstantial beliefs, different circumstantial beliefs will cause her to have different by-beliefs at different times, and the causation will be piecemeal. I specify the relation between the two events of having beliefs accordingly:

(d)(vi)(b) the agent's having circumstantial beliefs p-causes the agent's having by-beliefs.

In the second component of decision making, the agent passes from beliefs about how to move to reasons to move. So her by-beliefs should cause her to form subreasons and, again, the causation should be piecemeal: when things are working right, every subreason is caused by some by-belief, and different by-beliefs cause different subreasons at different times. So I specify the relation between her having beliefs about how to move and her having reasons to move in some of those ways as follows:

(d)(vi)(c) the agent's having by-beliefs p-causes the agent's having subreasons.

Movement modulation, the third activity that modulated movement involves, is easily characterized: the agent's reasons to modulate the total movement in certain ways cause her to do so. Again, the causation is piecemeal: different subreasons cause different submovements at different times, and every submovement is caused by some subreason. Thus I state the last clause of the definition as:

(d)(vi)(d) the agent's having subreasons p-causes the total movement e.

Finally, we consider the role of the guiding reason in the four instances of p-causation specified in clauses (vi)(a) through (d). The guiding reason guides in the sense that it provides the rationale for making the other instances of causation occur, and unifies the various parts of the various activities that make up modulated movement so that they are coherent and all contribute to the same end, satisfying the guiding desire. So the agent must have the guiding reason for as long as she carries out the total movement. Clause (b) of the definition of being intentional requires that the guiding reason cause the total movement, but clause (vi)(d) of the definition of causal normality requires that subreasons, rather than the guiding reason, cause the total movement. We can capture the idea that guiding reasons drive the component activities of modulated movement, and that reasons with more fine-grained content, subreasons, are what directly cause total movements, by taking the guiding reason to be a *second-order cause*: it is an event that causes other instances of causation to occur.⁶⁸ The notion of a second-order cause is familiar from daily life. Plugging in a radio is a second-order cause since it causes such things as pressing a button to turn on the radio; more idiomatically, we say that pressing the button turns the radio on if it is plugged in. In a similar way, in cases of modulated movement the guiding

reason is what causes the circumstances of the movement to cause the agent to have circumstantial beliefs, her having circumstantial beliefs to cause her to have by-beliefs, her having by-beliefs to cause her to have subreasons, and her having subreasons to cause submovements. So the sense in which guiding reasons cause total movements is that they create situations in which complexes of causally-related events cause total movements. Since the instances of causation that the guiding reason causes are piecemeal, and are spread out over time, the second-order causation by the guiding reason must be piecemeal too. In the definition I specify the causal role of the guiding reason in modulated movement as follows:

(d)(vi) as S carries out e, S's having the guiding reason p-causes (a) through (d).⁶⁹

I will now illustrate the analysis of causal deviance I have stated by showing how to use it to explain a paradigm case of action. Then, in the next section, I will apply the analysis to explain the difference between cases in which a person's movement of letting go of a rope is caused in a normal way, and cases in which the movement is caused in a deviant way, as in Davidson's example.

Suppose that Smith performs the action of reaching for a glass in order to pick it up and drink from it. Since Smith's reaching for the glass is her action, it is (by definition (A)) intentional for Smith under a description, which we suppose to be the description that Smith reaches for the glass. Smith's reaching for the glass is intentional for Smith under that description (by definition (I')) since, we further suppose:

There is a guiding reason to make it the case that Smith reaches for the glass such that:

- (a) Smith has the reason,
- (b) Smith's having the reason causes Smith's reaching for the glass,
- (c) Smith's reaching for the glass is a movement of Smith's body,
- (d) the causal chain from Smith's having the reason to Smith's reaching for the glass is normal.

Now let us apply the definition of modulated movement to explain clause (d) of the example. We suppose, first, that:

- (d)(i) Smith's reaching for the glass is a prolonged event.

We also suppose that:

- (d)(ii) Smith's having the guiding reason to reach for the glass is a prolonged event.

As I have suggested, Smith's having the guiding reason must persist through the interval over which she carries out the total movement. For suppose that it does not – suppose, for instance, that Smith forms the guiding reason, starts reaching for the glass, but changes her mind, and stops having a reason to reach for the glass. If she then continues to reach for the glass, she is doing something she has no reason to do, and so reaching for the glass is not (or, at least, is no longer) her action.

Next we suppose that when Smith makes the decision to reach for the glass, and forms a reason to do so, she initiates three processes of forming psychological states. First, she observes where the glass is and where her hand is, and, as she reaches for the glass, keeps observing the

glass and her hand, until she has grasped the glass with her hand. She makes these observations using different sensory modalities: she uses her sense of vision to see where the glass is, and her proprioceptive sense as well as her sense of vision to observe where her hand is. In virtue of these processes of observation,

(d)(iii) Smith has beliefs about the relevant features of her circumstances.

Second, Smith begins thinking about what she needs to do to satisfy her desire to reach for the glass. For instance, if when she begins to carry out her total movement her hand is on her lap, she may at that stage form and entertain the belief that she can satisfy her desire to reach for the glass by lifting her hand off her lap, and also the belief that she can do so by leaving her hand on her lap. At a later stage, if her hand is off her lap and at the same level as the glass, but her path to the glass is blocked by a pepper mill, she may form and entertain the belief that she can satisfy her desire to reach for the glass by moving her hand to the right, and also the belief that she can do so by moving her hand to the left. As a result of this thought process,

(d)(iv) Smith has beliefs about how to satisfy her guiding desire to reach for the glass.

Third, Smith begins a process of choosing to act on the basis of some of her beliefs about what she needs to do to satisfy her guiding desire to reach for the glass. That is, at each stage of reaching for the glass she settles on one of her beliefs about how to satisfy her guiding desire as the one to act on, forms a desire to act on it, and so - since she already has a belief that she can act on it - has a reason to act on it. Reaching for a glass is usually not a very complicated matter, and it may be obvious what she needs to do at each stage to reach for the glass. If so, she may form

just one belief about how to move to satisfy her guiding desire, and so will form a reason to move in that way without considering any alternative ways to move. Where what she needs to do is harder to determine – perhaps there are many other glasses on the table and she has to be careful she does not jostle them in reaching for her glass – she may need to form beliefs about other ways to move to satisfy her guiding desire, and form a reason to move in one of those ways after weighing the alternatives. But whether or not forming reasons to move in certain ways requires consideration of alternatives, there is the following process:

(d)(v) Smith has reasons about how to modulate her movement of reaching for the glass.

Finally, we consider the causal relations in which the processes specified in clauses (d)(i) through (v) stand. They are as follows:

(d)(vi) Smith's having a reason to reach for the glass p-causes:

- (a) the circumstances in which Smith reaches for the glass (which include her movement of reaching for the glass) p-cause her having beliefs about her circumstances,
- (b) her having beliefs about the circumstances in which she reaches for the glass p-causes her having beliefs about how to move in order to satisfy her guiding desire to reach for the glass,
- (c) her having beliefs about how to move in order to satisfy her guiding desire p-causes her having reasons to move in certain ways,
- (d) her having reasons to move in certain ways p-causes the total movement of Smith's reaching for the glass.

For instance, at the outset of carrying out her movement of reaching for the glass, the processes stand in causal relations like the following. The glass is on the table in front of her and her hands are on her lap. She forms a reason to reach for the glass, and observes where her hands are and where the glass is. She realizes that she can start satisfying her desire to reach for the glass by lifting her right hand off her lap, forms a reason to do so, and lifts her hand. Later, towards the end of the process of reaching for the glass, her fingers are right in front of the glass, but they are closed too tightly to grip it. She sees this, realizes that she can further reach for the glass by opening her hand, decides to do so, and does. At the end of the process of reaching for the glass, her hand is closed around it. She is now ready to pick it up and bring it to her lips, and initiates appropriate processes of belief formation about her circumstances, belief formation about how she can satisfy her (new) guiding desire of bringing the glass to her lips, reason formation about how to do so, and moving in accordance with her reasons.

2.3 Application of analysis

The example just described was intended to show how my analysis of being causally normal works by applying it to a paradigm case of action. My concern in this section is to show how to use the definition to explain why, in the example of the mountain climbers I described in the last chapter (in section 1.1.3), letting go of the rope was caused in a deviant way, and so was less intentional than if it had been caused in a normal way. I will first describe a set of conditions that, if satisfied, would make the causal chain from the agent's having a reason to let go of the rope to doing so normal. Then I will use that set of conditions as a basis for identifying various reasons why she might fail to modulate her movement properly, and the causal chain from her reason to her movement might be deviant.

We start by assuming that Smith, the climber holding the rope, lets go of the rope, and that her doing so is an action. We assume, further, that she lets go of the rope by opening her

hands enough to allow the rope to slip through them, and that her action is complete when the end of the rope has slipped through her hands. Since Smith's letting go of the rope is a full-blown action, it is (by definition (A))⁷⁰ intentional for Smith under a description. We suppose that one such description is a description of a bodily movement, that Smith lets go of the rope, and that the movement is intentional under that description (by definition (I'))⁷¹ since the following conditions are satisfied:

There is a guiding reason to let go of the rope such that:

- (a) Smith has the reason,
- (b) having the reason causes her to let go of the rope,
- (c) letting go of the rope is a movement of her body,
- (d) the causal chain from having the reason to letting go of the rope is normal.

Now we turn to the crucial part of the example, describing a set of conditions that, if satisfied, would make (d) true. This part is crucial since the contrasting example we considered in the last chapter, in which Smith lets go of the rope as a result of a deviant internal causal chain, like the present example, satisfies all the other parts of the definition of being an action, so it is to this part of the example that we must look to explain why there is no causal deviance in this example but there is in the earlier one. We suppose that when Smith begins the action of letting go of the rope, her hands are closed around the rope ten feet from one end of it; she opens her hands enough to let the rope slide through them; and she keeps her hands open until the end of the rope has slid through them. Accordingly,

- (d)(i) opening her hand to let go of the rope is a prolonged event.

We also suppose that

(d)(ii) having the guiding reason to let go of the rope is a prolonged event.

We suppose further that Smith has the following psychological states:

(d)(iii) she has beliefs about the circumstances in which she lets go of the rope,

(iv) she has beliefs about ways to move to satisfy her desire to let go of the rope,

(v) she has reasons to move in certain ways.

We suppose, as in the case in which letting go of the rope is caused in a deviant way, that when Smith forms her reason to let go of the rope, she has already observed that she holds the rope in her hands close to one end and that the other end of the rope supports her partner. Thus the process of belief formation about her circumstances does not have to begin in response to forming her guiding reason, though it must continue as she lets go of the rope for letting go to be normal. Perhaps her process of belief formation about ways to move her hands to let go of the rope also began before she formed the guiding reason to do so. Her mind may have wandered, and she may have pondered various results she could achieve by opening her hands - perhaps it was when she realized that she had a way to open her hands to let go of the rope that she formed the guiding reason to do so. But it was only after she formed her guiding reason that she began to consider in a systematic way how to move to open her hands to let go of the rope, so that she constantly revised her beliefs to reflect changes in how she was gripping the rope, and in other relevant features of her circumstances. Her process of subreason formation probably began only after she formed her guiding reason to let go of the rope. For most likely, if she did contemplate various

things she could do by opening her hands, she was not so reckless as to try any of them until she had made the decision that resulted in forming her guiding reason to let go of the rope (though the results of opening her hands are pretty obvious, and she didn't need to play with the rope to see that she could let go of it by opening her hands). However, in a situation with less at stake – say, in a beginning mountain climbing class where a safety rope also supported her partner – she might have played a bit with the rope she was holding to get a sense of how easy it would be to open her hands to let go of it. But as in the case of the process of by-belief formation, it was only after she formed her guiding reason to let go of the rope that she began to pass systematically from beliefs that she could let go of the rope by moving in a certain way to forming reasons to do so.

Since we are describing a case in which the causal chain from Smith's having a reason to let go of the rope to doing so is normal, we assume that once she has formed the guiding reason to let go of the rope, there are systematic relations among all the processes specified in (d)(i) through (v). Thus:

- (vi) Smith's having a reason to let go of the rope p-causes:
 - (a) the circumstances in which she lets go of the rope (including the movement itself) p-cause her having beliefs about her circumstances,
 - (b) her having beliefs about her circumstances p-causes her having beliefs about how she can move to satisfy her desire to let go of the rope,
 - (c) her having beliefs about how she can move to satisfy her desire to let go of the rope p-causes her having reasons to move in certain ways,
 - (d) her having reasons to move in certain ways p-causes her total movement of letting go of the rope.

Now let us use the conditions just described, which make the causal chain from letting go of the rope to doing so normal, to see how we can explain why the causation is deviant in the case in which the agent's having a reason for letting go of the rope causes her to panic, and her panic causes her to let go of the rope. As I said earlier, a movement is caused in a deviant way if and only if it satisfies all the requirements for being intentional in definition (I') except for (d), the requirement that the agent's having the reason for carrying out the movement causes the movement in a normal way. Thus the reason to make the movement must cause it. My analysis requires that a number of conditions be satisfied for a movement to be properly modulated and so normal, and a movement can be internally deviant because any of those conditions fails to be satisfied. It will be useful to start by distinguishing two ways in which a movement can fail to be normal: because there is a *complete* breakdown of the agency of the person whose movement it is, or because there is a *partial* breakdown of her agency. By a complete breakdown of a person's agency, I mean that she loses altogether at least one of the capacities required for her reasons to cause her movements in a normal way, and by a partial breakdown of a person's agency, I mean that at least one of the capacities is impaired. In cases in which her agency completely breaks down, she will have no responsibility for the movement she carries out, while in cases in which her agency is impaired, she may have some degree of responsibility for it, but less than if the movement were caused in a normal way.

Let us first consider why, as a result of a complete breakdown of a person's agency, letting go of the rope could be caused in a deviant way. We suppose that the mountain climber forms a reason to let go of the rope, panics, as a result of panicking loses entirely her ability to satisfy at least one of the conditions required for her reason to let go of the rope to cause her to do so in a normal way, and as a result of the loss, lets go of the rope in a deviant way. This can take a number of different forms. She may become so panic-stricken that her senses are dulled or lost – perhaps she suffers hysterical blindness and her hands go numb – and does not even realize she

has let go of the rope until after she has done so. In this case, she has lost altogether the capacity to form circumstantial beliefs. In a second case, she is keenly aware of the rope slipping from her grasp, but her panic makes her unable to think, and she cannot figure out what to do to keep from letting go of the rope until it is too late. Here she loses the capacity to form by-beliefs. If she could form by-beliefs, she could possibly figure out ways to move her hands to stop letting go of the rope, but she cannot, and lets go of the rope. In a third case, she is again keenly aware of her circumstances, and she thinks of various things she might do with her hands to keep from letting go of the rope. But in her panic she is unable to determine on which of them to act. Here she has lost the capacity to form subreasons. In a fourth case, she knows she is letting go of the rope, and knows exactly what she needs to do to keep from letting go, but is paralyzed by fear and cannot move in the ways she knows she must to hold on to the rope, and it slips from her grasp. Here she has lost the capacity to make submovements.

In the cases I have so far described, letting go of the rope is deviant because the agent loses the capacity to have at least one of the states modulated movement requires. In another set of cases, she retains the capacity to have the states, but in the case of at least one of them loses the capacity to make having the state stand in the appropriate causal relations. Perhaps her senses remain acute, but she cannot concentrate, and the relevant features of her circumstances do not cause her to form beliefs about them. She may retain the capacity to make decisions about how to modulate her movement, so that she forms beliefs about ways to move to keep from letting go of the rope, and she may also retain the capacity to form subreasons, but she may be unable to think straight, and the reasons may be unrelated to the beliefs. Or she may retain the capacity to move, but shake so much or be so distracted that her reasons to modulate her movements in certain ways do not cause her to do so. We can imagine an extreme case in which she can neither carry out any of the processes required for modulated movement nor make them stand in the proper causal relations: her panic causes her to faint, and fainting causes her to let go of the rope.

I turn now to consideration of why letting go of the rope could be caused in a deviant way as the result of a partial breakdown of the mountain climber's agency, a breakdown that leaves her capacity for agency impaired but not lost. I will consider just two cases, though there are many others. In both cases the climber forms a reason to let go of the rope, the reason causes her to initiate a modulated movement, and her awareness of her reason causes her to panic. In the first case, awareness of her reason to let go also causes her to lose her desire, and so her reason, to let go, and causes her to form a reason to hold on to the rope. But her panic causes a delay in stopping the old modulated movement and starting the new one, and as a result of the delay she lets go of the rope. In the second case, awareness of her reason to let go of the rope causes her to form a reason to hold on to it, but in her panic she also retains her reason to let go, so she is "of two minds" about letting go. As a result, the modulated movement of letting go continues – though in a defective form – and the rope slips through her hands. Both are cases in which letting go of the rope is causally deviant, but why?

In the first case, we can explain letting go of the rope as causally deviant because having a reason to do so *w-causes* the other instances of causation instead of *p-causing* them; that is, it only *sets off* the other instances and does not keep causing them as the total movement is carried out. We can explain the second case too as one in which having the reason does not p-cause the other instances of causation, but it is not because the causation is wholesale – having the reason to let go of the rope does not merely set them off, but to some extent drives them. On the other hand, if the climber is in a state of panic, and is of two minds about letting go of the rope, it seems plausible that having the reason causes the other processes in an intermittent or desultory way. Then, though having the reason may cause some parts of the other instances of causation, it does not cause *every* part of them – it does not cause them in a systematic way – as is required for p-causation. Call such causation *desultory* or *d-causation*.

How would the w-causation or d-causation, rather than the systematic p-causation, of the movement by the reason be reflected in letting go of the rope? We would expect the agent to lose her focus on letting go of the rope as she does so. In the case in which having the reason w-causes the other instances of causation, we would expect her to lose her focus once for all when she ceases having the reason to let go of the rope, and in the case in which having the reason d-causes the other instances of causation, we would expect her focus to keep wandering – she would be distracted as she lets go of the rope. It also seems plausible that losing her focus would affect the other instances of causation, so that they would not be piecemeal and systematic. For instance, she might fail to attend to her circumstances as closely as she would if she kept her goal of letting go of the rope clearly in mind, or might lose track of what features of her circumstances are relevant. And even if she somehow managed to keep her goal of letting go of the rope clearly in mind, it is likely her panic would affect the process of belief formation about her circumstances in ways like these. We would also expect there to be similar problems with her other states and their causal relations. If she loses her focus on letting go of the rope, or even if she retains her focus but is in a state of panic, she might form too many incorrect beliefs about how to modulate her movement, or form them only intermittently, or they might fail to be properly based on her beliefs about her circumstances. She might make the wrong choices about which beliefs to form reasons to act on, basing them on beliefs about relatively unpromising ways to modulate her movement, or she might form reasons to move in inconsistent or incompatible ways. Since letting go of a rope is so easy to do, she may well end up letting go of the rope even if she makes many mistakes, so there are many likely cases in which letting go of the rope will be deviant. But when what she is doing is more difficult – say, if she is trying to catch a fly ball, drive a car through traffic, or balance on one foot - if the capacities required for modulated movement are impaired, she is much more likely to fail to satisfy the guiding desire, rather than to satisfy it in a deviant way.

I will conclude my discussion of possible reasons why a movement can be deviant by describing two ways in which panic, or another strong emotion, might lead to impairment or loss of capacities required to carry out modulated movements. First, carrying out normal movements requires that the agent be able to form second-order beliefs – beliefs whose contents are propositions about the contents of other psychological states. To give just one example, for an agent to form a belief that she can satisfy her desire to let go of a rope by moving her hand in a certain way, she must be cognizant of what she desires, and must be able to form the belief that she desires to let go of the rope. We will say that an agent with the capacity to form second-order beliefs is *reflective*. It seems plausible that there will be cases in which an agent who is generally reflective will temporarily lose that capacity, and will believe or desire something without realizing she believes or desires it. One case in which an agent may temporarily lose the ability to be reflective is when she panics (though in cases of internal causal deviance, it is only because the agent was reflective before she panicked that she became aware of having the reason that caused her to panic). For instance, the climber may know that there is something she has a reason to do, and that to satisfy the desire she needs to be moving, but she may lose track of *why* she should be moving – of the content of her reason. Then she will be unable to form beliefs about specific ways to move in order to satisfy the desire that forms a part of her reason.

The second way I will describe in which a strong emotion can lead to impairment or loss of capacities required to carry out normal movements has to do with the need for an agent to form beliefs by drawing inferences from other beliefs. In the grip of a strong emotion, an agent may be unable to make such inferences quickly enough to move appropriately, or may be unable to make them at all. The mountain climber might observe that she is starting to let go of the rope, and also that her partner is starting to fall, but fail to infer that letting go of the rope is what is causing her partner to fall. In this sort of case, someone who knows that the agent has the two beliefs might be able to draw the inference on that basis, and in that sense the agent *carries or has* the

information that the conclusion expresses. Yet insofar as having the two beliefs does not enable her to act on that information, the information is not *available* to the agent – the agent’s panic may result in her inability to “put two and two together.”⁷² Due to a breakdown in her inferential system, being in the physical states underlying her beliefs that the premisses are true does not cause the physical state underlying a belief that the conclusion is true, and so does not cause her to act on the basis of the joint content of the two beliefs. This could be so because of the absence of a second-order cause: some physical state might fail to obtain that is required to cause the physical states underlying the beliefs about the premisses to cause the physical state underlying the belief about the conclusion.

2.4 Objections to the analysis

In the previous section I tried to show how my analysis of modulated movement can be applied to explain why some movements are internally normal and others are internally deviant. In this section I want to defend the analysis against three objections that might be raised against it: that many intentional movements are continuous, but the analysis explains deviance only in cases of intentional movements carried out in discrete steps; that some intentional movements are not modulated, and the analysis cannot explain deviant cases of those; and that the analysis fails because the causation of submovements by subreasons could be deviant. I will discuss each objection in turn.

2.4.1 Intentional movements can be continuous

One might think that the analysis presupposes that a movement can be normal only if it is carried out as a sequence of discrete steps. For instance, the analysis may suggest that in reaching for a glass, there would be series of events like the following: the agent forms a belief that her hand is six inches from the glass; she forms the belief that when her hand is six inches from the

glass she can contribute to satisfying her desire to reach for the glass by moving her hand one inch closer to it, and the belief that when her hand is six inches from the glass she can contribute to satisfying her desire by moving her hand one inch further from the glass; she forms a reason to move her hand one inch closer to the glass; and, finally, she moves her hand one inch closer to it. But this obviously does not capture what goes on in typical cases of reaching for a glass. As I have described the example, the movement of reaching for the glass would be carried out in a series of discrete steps, with the agent moving a bit, stopping while she assesses her progress and plans out the next stage of the total movement, then moving a bit more, and so on. When what a person is doing is very difficult or complicated, or where something unexpected comes up, she may indeed pause as she carries out a total movement so she can think. But it is clear that in typical cases in which a person does something relatively simple, like reach for a glass, she does not stop and start her total movement.

One possible explanation of why there are no pauses in such cases is that when the agent is ready to carry out a certain step of reaching for the glass, she has already planned it out, since as she carries out each part of the total movement, she concurrently determines how to carry out the next. The problem with this explanation is that it still seems to fail to capture what goes on when a person intentionally reaches for a glass. We do not ordinarily conceive of the total movement as divided into shorter segments that each takes place over an interval of time. Instead, carrying out the total movement seems to be a continuous process. Further, the process of obtaining feedback about the total movement – that is, of forming circumstantial beliefs – also seems to be a continuous process. A person might be doing two things at once, such as reaching for a glass while she is watching television, so that she switches her attention back and forth between the television and her movement. In this case, receiving feedback about the relevant circumstances of her movement would not be a continuous process. It is clear, however, that this case is different from many cases of reaching for a glass, in which the agent keeps observing the

circumstances of the movement until it is complete, so that the process of forming beliefs about the circumstances seems continuous. But if we assume that the process of forming beliefs about the circumstances is continuous, another problem arises. The agent would have to form an infinite number of beliefs in a finite amount of time, but it may not be clear this is possible. If having a belief is a matter of standing in a certain relation to a sentence, or sentence-like object, formulated in the language of thought, how could the agent form an infinite number of beliefs in a finite period of time? There could be an array of physical states each corresponding to a sentence with a slightly different content – for instance, the index finger of her hand is now one-hundredth inch closer to the glass, now another one-hundredth inch closer and so on - and as the agent observes her movement, the physical relation to each physical sentence-like object constantly changes; it would be as though she keeps moving her finger to point to a different sentence in a list of sentences printed on a page. But how could there be an infinite number of such sentences? And even if there could be an infinite number, how could an array of sentence-like objects represent the different positions reached in carrying out a continuous movement, since the sentences could represent only discrete facts? It is possible that the different positions represented are so close together that the agent cannot tell the difference between a case in which she perceives a movement that is in fact continuous and a case in which the movement proceeds in imperceptibly short discrete steps, like a movie. Still, this explanation does not explain, but explains away the intuition that the process of perception of a continuous movement is itself a continuous process, and it would be better on theoretical grounds if our theory could preserve the intuition. Similar problems arise for the formation of by-beliefs and subreasons (and so for the formation of desires).

Another problem I will mention is that my analysis of causal deviance seems to require in many cases that an agent make complex decisions – that is, form by-beliefs and subreasons - about how to move too quickly to be plausible. As I have suggested, how to modulate a

movement in some cases, such as typical cases of reaching for a glass, is obvious, and alternative ways to modulate it need not be considered. But what about in more complicated cases, such as throwing a football to a receiver running at full speed? Slight differences in the way in which the quarterback's arm moves, and the timing of the release of the ball from her hand, can make substantial differences in how easily, and whether, the receiver can catch the ball. If we were to program a computer to simulate the action of the quarterback, it seems reasonable that we would have the computer carry out a large number of complex computations to determine the best way, or at least a satisfactory way, to move her arm and release the ball. We are certainly not aware of carrying out such processes of decision making when we throw a ball. But even if we assume that such decision making is carried out unconsciously, it is still not apparent how a person could carry out all these computations in the short time she has to make her decisions.

I think we can see how to solve these problems if we consider some recent research regarding the neurobiology of movement control.⁷³ This research suggests that the organization of the neural systems that control movement is hierarchical and modular. For instance, isolated parts of the body, such as the feet, are themselves able to modulate their movements in response to sensory input they receive.⁷⁴ Further, the neural systems responsible for reflexes appear to be the same as those activated in voluntary movements.⁷⁵ Thus there is evidence that suggests a view of the human neural system on which the capabilities of acquiring and processing information about the course of movements, and modifying the movements in response to the information, are distributed throughout the body, and the capabilities are exercised in reflexes as well as in voluntary movements. On this view, when the hiker in our earlier example makes a conscious decision to take a step, what happens would be something like the following: A signal from her brain activates neurons in her spine, which activate neurons in her legs and arms. These neurons in turn activate neurons that control the movement of different parts of her legs and arms, resulting in the lifting and planting of her feet, and the swinging of her arms, on the basis of

locally available information; for instance, it is known that applying pressure to different toes can initiate different reflexes that control the movements of individual toes in a way functionally appropriate for maintaining posture.⁷⁶ Taking this view leads to a picture of human actions as naturally decomposable into actions, or action-like events, of parts of the body. For instance, when a person reaches for a glass to pick it up and drink from it, there might be a natural decomposition into four different parts – four separate reflexes (reaching for the glass, picking it up, bringing it to her lips, and drinking from it) - and the action as a whole and its components will all be modulated movements. Higher levels of the action will control the action in response to less local information, and lower levels will do so in response to more local information. For instance, the hiker may see before she takes a particular step that she will have to jump over a rock, and in taking the step consciously modify how she plants her foot to jump off it.

To illustrate modulated movement as it might occur in the foot, and to show how we can view even low-level movements under peripheral control as involving processes of belief and desire formation, I will consider the example of planting a foot in more detail. Imagine the following scenario. The foot descends towards the ground, and the toes touch the ground first. There is a set of neurons in the foot that are sensitive to pressure on the sole, and another set of neurons in the foot that are sensitive to internal stresses that result from the foot being flexed or extended to various degrees. There is a third set of neurons that respond to the activity of the other two sets of neurons by firing in certain ways, and these neurons provide input to motoneurons in the foot that activate and inhibit specific muscles in the foot depending on the input they receive from the third set of neurons. How the motoneurons activate and inhibit the foot muscles is systematically related to how the neurons from which they receive input fire, and as a result of the systematic relation the foot moves in ways functionally appropriate to plant itself and maintain the hiker's posture.

The pattern of firing of the neurons sensitive to pressure on the sole of the foot is counterfactually dependent on the pattern of the pressure, and so the pattern of firing *indicates* what the pattern of pressure is.⁷⁷ Similarly, the pattern of firing of the neurons sensitive to internal stresses indicates what the pattern of internal stresses is. In addition, the pattern of firing of the third set of neurons indicates the conjunction of what the first two sets of neurons indicate about the foot: it indicates what the pattern of pressure on the sole is *and* what the pattern of internal stresses is.⁷⁸ On a certain naturalistic conception of belief, a conception developed and defended by Daniel Dennett⁷⁹ and Robert Stalnaker⁸⁰ among others, the pattern of firing of the third set of neurons is a form of belief, since it is an indicating state that disposes the agent to act in ways that would tend to satisfy her desires if it, and her other beliefs, were true.⁸¹ Since the motoneurons respond to the firing of the third set of neurons by causing certain movements in the foot, on the conception of desire correlative with this conception of belief, the pattern of firing of the motoneurons is a form of desire, in this case a desire that she make those movements, since it is a state that disposes an agent to act in ways that would tend to bring about her making those movements if her beliefs were true.⁸² Thus, on this conception of belief and desire, there is nothing problematic about forming an infinite number of beliefs and desires in a finite period of time, or about processes of belief and desire formation that are continuous. So the processes of propositional attitude formation that my analysis of modulated movement requires are consistent with the existence of a continuous feedback mechanism.

So far I have argued that there is a sense in which the foot engages in processes of belief and desire formation, and that these processes can be continuous. In the example of the foot, the beliefs formed are beliefs about the course of its movement and other features of its circumstances – circumstantial beliefs – and the desires are desires that the foot move in certain ways – desires that are constituents of subreasons. But my analysis of modulated movement also requires that for a movement to be modulated, the agent – in this case, the foot – form beliefs

about ways to move to satisfy the guiding desire – by-beliefs – and evaluate them to produce subreasons. As I have mentioned, though in many cases what needs to be done is obvious, and little work will be involved in determining possible ways to move and settling on one of them, in many other cases, such as throwing a ball at a moving target, or walking over a very uneven surface, the decision making process that yields subreasons will be much more involved, since more alternative ways of moving will need to be considered, and they will need to be considered in greater detail.

There are at least two ways in which an agent can speed up the production of by-beliefs and subreasons. The first is to carry out the decision making process before the modulated movement begins. This can take several forms. The way an agent modulates a movement might be determined by evolution, as in the case of the innate way the foot modulates its movement when it is being planted. Here the process of by-belief formation and evaluation is carried out by a succession of ancestors of the agent. Alternatively, the way in which an agent modulates a particular movement might be the result of training or practice. A person playing a game of basketball might know instinctively how to move her arms and hands in order to sink a basket from a particular spot on the floor as a result of long hours spent trying different ways of doing so. A third way is to plan how to make a movement before beginning the movement. The basketball player might plan during a time-out how to modulate a movement to shoot the ball, exploring different ways of doing so and settling on one, and then, when play resumes, execute the modulation when the situation is appropriate.

The second mechanism that can be used to speed up the decision making process is to constrain the ways the agent can move, eliminating certain unpromising options, by providing the agent with a specific structure that allows it to move in some ways but not others. In the case of agents that are living organisms, evolution provides the agent with this structure. For instance, when a person reaches for a glass and brings it to her lips, the ways in which she can do so are

constrained by the relative lengths of her arm above and below the elbow, and by the structure of her shoulders, elbow and hand, which have only a few degrees of freedom in which they can move. As a result, there is a limited range within which someone bringing a glass to her lips can hold the glass relative to her body. If the right elbow is bent so the forearm and upper arm form an angle of less than twenty degrees, structural constraints force her to bring the glass to a position somewhere between mid-chest level and a few inches to the right of her right ear. We might attribute to her the belief that she can satisfy her desire to bring the glass to her lips by holding it with her right hand somewhere between mid-chest level and just to the right of her ear, and the belief that holding the glass below waist level is not a way in which she can bring it to her lips. But, typically, these attributions are warranted not because she has explicitly considered both possibilities, and has accepted the first and rejected the second. Instead, they are warranted because the structure of her body limits her options, allowing her to move within a range of relatively useful ways, and preventing her from moving in a much larger number of relatively useless ways. As a result, the decisions she has to make about how to bring the glass to her lips are limited to considering different possibilities within the relatively narrow range of ways in which her structure allows her to move.⁸³

I have argued that modulated movement is a pervasive phenomenon, a phenomenon that occurs not only at the level of complex actions that require conscious thought, but also at the level of movements carried out with little or no thought. I have done so by arguing that even in the case of movements carried out with little or no thought, the agent, or parts of her body, form the beliefs and desires required for modulated movement. A critic might object that my argument depends on an excessively broad notion of belief and desire, a notion that one who is not sympathetic to a particular sort of naturalistic account of belief and desire will reject. For we ordinarily think of beliefs and desires as psychological states, which are states of the brain or central nervous system, not states of the feet, arms or hands.

We can respond to the critic by pointing out that even if it is controversial whether states of the feet can be beliefs and desires as we ordinarily think of them, it is obvious that the feet have states that carry information and states that bring things about, and that these states are causally related to the circumstances of the agent and her movements in such a way that the movements are functionally appropriate and useful given the circumstances. So even if the states are not strictly beliefs or desires, they bear important similarities to belief and desire, and can be viewed as the same general kinds of things. Also, though the naturalistic conception of belief and desire I have described assumes there are important similarities between a wide variety of information bearing states and bringing-about states, it also assumes there are different kinds of information bearing states and bringing-about states, and important differences among the kinds. For example, we might draw a distinction between two kinds of beliefs that have the content that an agent cannot move her hand below waist level if she bends her elbow to form an angle of less than twenty degrees. One kind of belief we attribute to her because her arm has a certain structure, the other because we see her observing how she can move her hand with her elbow bent. We attribute the first belief because she is *in* the state that the proposition that forms the content of the state implies her to be in, while we attribute the second belief because she is in a state that *explicitly represents* her as being in the state that the proposition implies her to be in, a state distinct from the state the proposition implies her to be in. There is a sense in which the first belief is *self*-representing while the second belief is *other*-representing, and we might require that for a belief state to be a belief as we ordinarily think of it, it must be other-representing.⁸⁴

2.4.2 Ballistic movements

I have just defended my analysis of internal causal deviance against the objection that it applies only to movements that are carried out in discrete steps. I will now consider a set of cases that may seem to pose another problem for my analysis. On the analysis, movements caused in

normal ways can be distinguished from movements caused in deviant ways by considering whether the movements are properly modulated. Now modulating a movement requires that the movement take place over an interval of time of a certain minimum length, an interval long enough for the agent to acquire information about the circumstances of the movement and to respond to the information. But one might object that not all movements are like that. It is obvious that an agent has time to receive and respond to feedback in many cases, such as catching a ball, steering a car, or reaching for a glass, but it is not obvious that an agent has time to do so in many other cases, such as pushing a button or pulling a trigger. Let us call actions that involve quick movements like pushing a button *one-shot actions*. The objection amounts to the contention that one-shot actions involve, not modulated movement, but ballistic movement, movement that is w-caused and not p-caused (or d-caused) by the guiding reason. Brief movements can only be w-caused, so all brief movements are ballistic, as well as all prolonged movements that are w-caused by a guiding reason. Since my analysis requires that a movement be p-caused to be normal, and since all ballistic movements are w-caused, the existence of deviant ballistic movements, brief or prolonged, would pose a problem for my analysis, since deviant ballistic movements could not be explained as deviant because they were defective modulated movements. I will respond to the problem that ballistic movements raise by first arguing that several cases of intentional movements that are apparently ballistic, despite their appearance, actually do involve modulated movement, and then arguing that in the case of movements that really are ballistic, the distinction between normal and deviant movements does not apply.

Let us start by considering an action of pressing a button that turns on a light. When a person does so, typically, she must first move her arm and hand so that one of her fingers comes into contact with the button. This movement, however, is clearly a modulated movement, so in such cases the assumption that pressing a button is ballistic is false. We can imagine a case in which a person discovers that her finger happens to be resting on a button, so moving her arm and

hand to the button is not part of her intentional movement of pressing the button. Suppose, further, that the button is on a hair-trigger, so that an extremely slight increase in the pressure of her finger on the button will depress it enough to turn on the light; that is, we suppose that the sensitivity of the button makes it a practical impossibility for the agent to respond to the course of her finger's movement in pressing the button before the button has been sufficiently depressed to turn on the light. But even in this case, though the event of pressing the button far enough to turn on the light may be brief, the person's movement is not: she will press the button until she *perceives* that it has been pressed in far enough to turn on the light, or sees that the light is on. Similarly, even when a person performs the one-shot action of pulling a trigger to shoot a gun, we draw a distinction between cases in which a person *squeezes* the trigger and cases in which she *jerks* the trigger. In cases like the first, she seems to have more control over shooting the gun, and is more responsible for doing so, than in cases like the second. Considerations like these show, I think, that virtually all intentional movements are prolonged and modulated rather than brief and ballistic.

Another kind of case that might seem to pose a problem for my claim that one can distinguish normal from deviant movements by considering whether the movements are properly modulated are cases in which a person intentionally punches, hits, kicks, or slaps something, as when a boxer throws a punch in a fight, or a soccer player kicks a ball. These might seem to be cases of action in which the causation of the movement of an external object by a movement of the agent's body is wholesale, and the movement of the agent's body is brief and so ballistic.

It is not clear to me that even in these cases the causation of the movement of the external object must be wholesale and without significant structure, nor that cases in which the causation is wholesale are common, but let us grant that there are some such cases of action that must be explained. To do so, the crucial move is to distinguish what is required for the agent to be responsible for the movement of an external object from what is required for her to be responsible

for a movement of her body. My claim is that for an agent to be responsible for the movement of an external object when the movement is w-caused by a movement of her body, she must be responsible for the movement of her body, and, for her to be responsible for that, the movement of her body must be a modulated movement, p-caused by her having a reason for the movement.⁸⁵

One reason to think that an agent is responsible for a movement of her body only if it is prolonged and modulated is that we ordinarily distinguish cases of, say, punching a person or kicking a ball in which the agent “follows through” from cases in which she does not. It seems clear that when the agent does not follow through, she has less control over, and is less responsible for, the movement she makes, and for its effect on the person or the ball, as when she does. Another, related reason is that there seems to be a clear difference between a case in which, say, a soccer player starts to kick a ball, and becomes distracted and looks away as she kicks it, and a case in which she attends closely to the entire movement. When an agent looks away, she seems clearly to have less control over the effect of her kick on the ball, and over her kick itself, than if she attends closely to her entire movement, and her responsibility for her movement and its effect on the ball seems diminished or absent. The last reason I will give is that, though we can imagine a person throwing a punch by throwing her hand as she might throw a ball, so that once she starts to move her hand she becomes passive and relies solely on her hand’s momentum to complete the punch, this is clearly not what one does when one intentionally throws a punch. For one thing, when an agent throws a punch, she keeps her hand closed to form a fist. She does not begin the movement and then go limp; even as she throws the punch, she still has things to do. However, if these movements really are brief and so ballistic, it should not matter whether an agent follows through, continues paying attention to the movement as she makes it, or goes limp after she has started the movement, since an agent’s having reasons would w-cause her movements, and her psychological role in the movements would be confined to setting them off and would end once they begin. Thus even in cases of throwing a punch or kicking a ball, and

even if we grant that in those cases an agent's body collides with an external object in a way analogous to the way in which one billiard ball collides with another billiard ball, it is clear, I think, that it is more plausible that the causation of the movement of the agent's body by her having a reason for the movement is piecemeal rather than wholesale, and that the action is modulated rather than ballistic.

Consider another case in which an agent might seem to carry out an intentional movement that is brief and so ballistic. Suppose that Jones twitches intentionally when enemy agents pass as a signal to a gunman to shoot them. Is a twitch really a movement that is brief, and so ballistic? I don't think it is. Even a twitch has a complex structure that a collision of billiard balls lacks, and even in the case of an involuntary twitch it seems likely (as I discussed in the previous section) that the muscles involved in making the twitch receive and respond to feedback about the course of the twitch. But suppose for the sake of argument that the twitch in our example really is brief and ballistic. What would be required for twitching when an enemy passes to be intentional? Intuitively, if twitching is intentional, the agent must twitch only in appropriate circumstances – when and only when an enemy is passing. But this means a ballistic twitch can be intentional only if it is part of a modulated movement. Another way to make this point is to consider what conditions must be satisfied for a ballistic twitch to be deviant. Suppose Jones forms a momentary desire to twitch when an ally goes by, becomes aware of her reason, panics, and twitches because of her panic. What makes this twitch deviant is not the mere fact that Jones twitched, but that she twitched *when* she did – when an ally and not an enemy passed. So a ballistic twitch can be deviant only if it is part of a prolonged movement that, if normal, would be made to depend systematically on the agent's circumstances, and so would be a modulated movement. Of course, most of what we are calling the movement in this case is not strictly a movement, but the absence of one. So here we can follow Davidson and construe movement broadly enough to cover being still,⁸⁶ or we can replace talk of movement with talk of changing

or maintaining one's position. Unlike more typical modulated movements, such as reaching for a glass or walking, an intentional ballistic twitch involves only two relevant ways of moving, and neither way of moving is required to prepare the ground for, or enable, the other way of moving.

I have just argued that intentionally twitching when a particular person goes by is part of a modulated movement. One might object that it is part of a modulated movement only because it must be carried out in specific circumstances to be carried out successfully. What about a twitch that is intentionally made, but in any circumstances the agent likes? Suppose that Brown forms a desire to twitch at some point in her life, and one day twitches. A critic might claim that Brown need not have monitored her circumstances to determine when to twitch, so the twitch was not part of a modulated movement. Is this twitch an example of an intentional movement that is brief and so ballistic?

My response is to claim that even in this case Brown must monitor her circumstances to twitch intentionally. Suppose Brown has decided to satisfy her desire to twitch once in her life by twitching her finger, and she has decided the time is now. To satisfy her desire, she must pursue one of two strategies, which both require receiving and responding to feedback about her circumstances. Either Brown tries to twitch, observes whether she succeeds, and if she fails tries again. Or she observes her circumstances, determines that they are suitable for twitching, and twitches. So this twitch too must be part of a modulated movement to be intentional. What this example shows is that even in the case of a movement as subtle as a twitch, there are circumstances in which a person could fail to carry it out successfully - something could be in the way of her finger and block its movement. And this will be true of any possible movement, including being still: there are circumstances that can make carrying it out impossible.

The examples of intentional twitches I have given show that when we speak of modulated movement, we should construe movement broadly enough to cover any case of intentionally changing or maintaining one's position. What about a case in which a person intentionally

collapses on the floor? Here the person is neither intentionally maintaining nor changing her position, but intentionally being passive; collapsing does not seem to be an exercise of one's agency, but a suspension of it. Intentionally collapsing seems to be a ballistic movement (though obviously a prolonged one), since one's reason to collapse seems to w-cause one's suspension of agency. But a closer look shows that intentionally collapsing should, after all, be assimilated to modulated movement: when an agent collapses intentionally on the floor, she must receive feedback about her circumstances – about whether or not she is flat on the floor – and respond to the feedback by continuing to be passive until she is flat on the floor. To suspend intentionally one's ability to change or maintain the position of one's body and let gravity take over is not after all to suspend one's agency, and should be regarded as a modulated movement.

I have been defending my analysis of normal movement as properly modulated movement against the objection that some intentional movements are ballistic, and my analysis cannot be used to explain normal cases of these. So far I have argued that several putative cases of intentional ballistic movements are really cases of intentional modulated movements. Thus the force of my argument so far is that if there are intentional ballistic movements, they are an unusual case of action, so my analysis of internal causal deviance still applies to most cases of action. However, I have not shown that intentional ballistic movements are impossible, and it is left open as a theoretical possibility that there could be deviant ballistic movements that would fall outside the scope of my analysis. I will now consider a genuine ballistic movement, and will argue that it is not the kind of movement to which the distinction between internally deviant and internally normal applies. The example of collapsing shows that for a movement to be ballistic, for it to be w-caused by a guiding reason to make the movement, it is not enough that we would ordinarily say that the agent initiates the movement but fails to change how or whether she moves as she carries it out. What is required is that the agent be *unable* to change how or whether she moves, so for a movement to be ballistic it must be *impossible* for the agent to make any relevant

responses on the basis of feedback she receives about it. For instance, for collapsing to be ballistic, once an agent has started her collapse it must be impossible for her to choose whether to collapse.

Consider an intentional movement like this: a person jumps off a ledge. It seems clear that this movement could be deviant, since the person could form a momentary reason to jump, panic when she realizes what she is thinking, and jump because of her panic. Have we found a deviant ballistic movement, a deviant movement my analysis cannot explain? I don't think we have. When a person intentionally jumps off a ledge, her action naturally divides into two parts: stepping off the ledge, and falling to the ground. Intuitively, the falling is intentional, and is a ballistic movement, since the agent initiates the falling but has no control over whether she falls, or how she falls, once she has stepped off the ledge (or, at least, she has no control over how she falls in any relevant way, since she lacks the needed aerodynamic properties). However, though the *falling* is ballistic, the *stepping* is obviously modulated, and it seems clear that the falling is intentional only because it is caused by a modulated movement of the agent. For if what caused her to fall from the ledge was a gust of wind, then the falling would not be intentional. Thus while the stepping is an *action* and is intentional in the strict sense I defined earlier,⁸⁷ the falling is an *effect* of an action and is intentional only in a derivative sense, as an effect.⁸⁸ One might object that if the person on the ledge asks someone else to push her, then the falling is intentional, but it is not caused by her modulated movement. Here it is clear that falling is something she *desires*, but it is not clear that it is *intentional*. But let us suppose that it is. Even in this case, there is still a relevant modulated movement of the person who falls: her asking plays a causal role in her falling. These considerations show, I think, that genuine ballistic movements that seem deviant only seem so: they are deviant only in the sense that they are caused by deviant modulated movements.

2.4.3 *Deviant submovements*

I have defended my analysis of modulated movement and normal causation against two objections, that some intentional movements are continuous, and the analysis cannot explain them; and that some intentional movements are ballistic, and the analysis cannot explain them. The last objection I will consider goes as follows. The analysis is intended to specify conditions that guarantee that a movement is properly modulated and its causation by a reason is normal. Among these conditions is the condition that subreasons cause submovements. But then the analysis is circular: cases in which the causal chain from subreasons to submovements is normal must be distinguished from cases in which the causal chain is deviant.

This objection is based on a misunderstanding of the analysis. The analysis does not require that subreasons cause submovements normally, only that subreasons cause submovements. So the analysis is not formally circular. Still, there is a related objection that might be made. Suppose a person performs a modulated movement, and the conditions required for it to be normal are satisfied. She acknowledges that she formed all the relevant psychological states, and that they stood in the causal relations required by the analysis. But the agent disavows responsibility for the movement on the grounds that each time she formed a subreason, she panicked when she realized she had, and panic caused each submovement. Should we accept her disavowal? That is, does the example show that the analysis could work only if it *were* circular, and required that the causation of each submovement by the corresponding subreason were normal? Such circularity would, however, be problematic, for the recursion cannot be continued indefinitely. Consider our hiker, who traverses the piece of rugged terrain. We might explain the traversal as normal because it can be divided into smaller intentional parts, say, the steps she takes, and we might explain each step as normal because it can be further divided into normal submovements. But we cannot continue this recursion indefinitely, for at some point we would have to explain such things as why causal chains from the firing of individual neurons to the

contraction of individual muscle cells are normal, and we cannot explain those as normal in the same way because we have run out of relevant levels of description.

One response is to deny that the recursion must proceed through different levels of description. For instance, we might explain dragging a finger in a line for ten centimeters as intentional because it is made up of ten intentional submovements of dragging the finger over one centimeter intervals, each of those is made up of ten intentional submovements of dragging the finger over ten one millimeter subintervals, and so on. I have already argued that a person can form an infinite number of beliefs and desires, or at least belief- and desire-like states, in a finite period of time, so what we might require for dragging the finger in a line for ten centimeters to be normal is that there be an infinite number of submovements with the requisite etiology, or perhaps that they be continuous. The critic might still object that even if the submovements are continuous and are all outcomes of the processes specified in the analysis, it is still a conceptual possibility that the causation of each submovement by the corresponding subreason is deviant. However, I don't think it is obvious that this *is* a conceptual possibility. On a roughly pragmatic view of agency – a view Davidson seems to take⁸⁹ - the existence of a systematic connection between a movement an agent makes and the circumstances in which she makes it, as a result of which the movement furthers such ends as her survival, well-being, or pleasure, is *constitutive* of agency, and provides all the grounds needed for attribution of the relevant beliefs and desires.⁹⁰ A theorist who takes this general view might still accept an agent's disavowal of responsibility for the submovements on the grounds that she did not really want to carry them out, and her agency was defective. But it would not be defective because of causal deviance, but because of weakness of will: the person's agency was not properly integrated, and we might think of this case as one in which multiple agents with conflicting goals inhabit her body.⁹¹

There is a second way to deal with the problem posed by deviant submovements. The strategy is to agree that the analysis should be applied recursively across different levels of

description, but only down to a certain level. Then, at that level, a different set of criteria is used to decide whether the causal chains from subreasons to submovements are normal – so at this level, normal movements are not explained as properly modulated movements. Recall that when we applied the analysis recursively to explain the hiker’s action of traversing a piece of ground, we reached a level of description at which the causation of submovements by subreasons is a matter of the causation of contractions of individual muscle cells of the foot by the firing of individual neurons. How could we decide whether individual neurons cause individual muscle cells to contract in a deviant way? At this level of description, the answer seems clear. The neurons and muscle cells have a particular organization and structure, and we can tell how individual neurons must cause individual muscle cells to contract to be working in a normal way. If they are not working as they normally do - say, if a neuron is damaged and no longer has the right connections to other neurons, so its firing causes a toe to move slightly up instead of slightly down – and if its firing nevertheless leads to the toe moving slightly down – say, if the toe’s going slightly up then causes another damaged neuron to fire, and its firing makes the toe go slightly down – then the causal chain from the firing of the first neuron to the toe’s going down is deviant. The reason why there is a problem of internal causal deviance in the first place, and why it seems so daunting, is, I think, that sorting out which of an extremely large number of possible physical paths from having a reason to making a movement are the normal ones seems to be an impossible task. But at the level of description of the firing of specific neurons and the contraction of specific muscle cells, how to sort out which possible causal paths from one to the other are normal is a manageable task with a clearly determinate answer.

2.5 Conclusion

I have argued that it is possible to say in physicalistic terms how causal chains from reasons to movements that are normal are different from causal chains that are deviant. If I am

right, then the problem of internal causal deviance is not an insurmountable problem, a problem that spells doom for any causal theory of action whose aim is to give a physicalistic reduction of action. But if something like the solution to the problem of internal causal deviance I have proposed is right, why have theorists such as Davidson been so pessimistic about solving it? I think there are two main assumptions these theorists may make that could explain their pessimism. The first assumption is that the causation of an intentional movement by an agent's reason must be instantaneous or wholesale. This assumption would lead one to overlook the crucial role that feedback, and modulation of movement based on feedback, play in typical actions, and would explain why these theorists overlook the possibility of modulated movement, as I discussed in section 1.4.1 of the previous chapter. The second assumption is that the obvious complexity of human beings means that it must be impossible to give systematic explanations of their behavior in simpler terms. Theorists who make this assumption do so, however, in the face of considerable recent evidence, both empirical and a priori, that complex organisms are highly structured, and have a modular, hierarchical organization. In virtue of this organization the activities of acquiring information about the world, processing this information, and moving in response to it are distributed throughout the body and occur at different levels of description. On my view, as a result of the distributed character of action, it can be understood in physicalistic terms.

*Chapter 3**Indication**3.1 Introduction*

In the last chapter I developed a theory of action – more specifically, a theory of rationalization of action – with two aims. First, the theory is based on Davidson’s theory of action, which provides an account of the difference between central cases of action and mere happenings, and I wanted to preserve this feature of his theory. Second, I wanted to extend Davidson’s theory to solve a problem his cannot solve, the problem of internal causal deviance. For the theory of rationalization I developed to be part of a general physicalistic theory of action, it must in addition provide a physicalistic reduction of the psychological states that figure in the theory of rationalization. In this chapter I want to focus on one of these states, belief. Following a number of philosophers,⁹² I will view belief as a species of *indication*, the relation that holds, for instance, between the fact that a thermometer has a certain reading and the proposition that the air surrounding it has a corresponding temperature. Indication is not all there is to belief: for a state to be a belief, it must both indicate something *and* cause the agent to behave in a way that satisfies her desires if it and her other beliefs are true.⁹³ But indication is much of what there is to belief.

My aim in this chapter is to develop a theory of indication that provides a physicalistic reduction of it. A complete physicalistic reduction of action would have to provide a physicalistic reduction of belief more generally, as well as of desire, the psychological states that figure in the theory of rationalization I proposed. Though I do not provide such reductions here, in giving a physicalistic reduction of indication I hope to go some distance towards giving a physicalistic reduction of belief, and to suggest what detailed physicalistic reductions of belief and desire might look like. In section 3.2 I propose and defend a theory of indication whose aim is to explain

what it is for an object or system to indicate a proposition by being in a single state. In section 3.3, I extend the theory to show how indicators can be combined to indicate more specific or precise propositions than any of them can indicate alone. In this way, I hope to show how it is possible for a simple mechanism to indicate a proposition of the degree of specificity and precision we ordinarily associate only with people. I conclude in section 3.4 with some brief remarks suggesting how the theory of indication developed in this chapter can be extended to explain indication in people. Before I turn to the theory, however, I want to make a few remarks about the methodology I will follow in developing it.

In developing my theory of indication I will not directly consider indication in people, but, instead, in simple mechanisms. But if my theory of indication is intended to form part of a theory of *human* action, why consider indication in simple mechanisms? Why not follow the more traditional philosophical method of linguistic analysis, of giving a systematic account of the language we ordinarily use to talk about indication and belief in people? There are three reasons.⁹⁴ First, one aim of the theory of indication I will develop, and, more generally, of the theory of action of which it is a part, is to help us see and understand the commonalities as well as the differences between people and the rest of the world. I want to show how human behavior could be part of the natural world by showing how we can see it as an incremental outgrowth of much simpler phenomena. Accordingly, the strategy I propose is to explain human psychological states and action by giving a general theory of psychological states and action that shows them to be widespread phenomena, and that explains psychological states and action in people as special cases.

It is important to see that I am *not* claiming that indication in the simple cases I will consider is just the same as indication in human beings. That there are important differences between a thermometer's indicating that a room is a certain temperature and a person's doing so is clear, and I will take the existence of such differences to be a datum of the theory I will develop.

But the fact that it is natural and common to extend talk of psychological states and action to much simpler, non-human things, and to apply the same distinctions, shows that we ordinarily recognize similarities between, on the one hand, the psychological states and the intelligent, rational behavior of people, and, on the other hand, the states and behavior or operation of simpler systems. It is only because we recognize such similarities that it even makes sense to talk of simple analogs of human psychological states and action, or, like John Searle, to criticize imputing intentionality to simple mechanisms because such intentionality is merely "as-if."⁹⁵

Thus the method I will follow bears more similarity to the traditional method of linguistic analysis than it may at first appear. There are, of course, ordinary concepts of human psychological states and action, and our philosophical theories of psychological states and action should explain them. But there are also ordinary concepts that are extensions of those concepts that we apply to analogs of human psychological states and action, and I take these extended concepts too to be concepts our theory of psychological states and action should explain. Both the method I will follow and linguistic analysis aim to make sense of ordinary talk, but the talk the method I will follow aims to explain has a wider scope than the talk traditional linguistic analysis aims to explain, encompassing not just talk of human psychological states and action, but also talk of simple analogs of human psychological states and action.

The second reason for considering indication in simple mechanisms rather than in people is heuristic. It is easier to understand the operation of simple systems than the operation of more complex ones - simple systems have fewer parts as well as fewer extraneous features, and it is easier to see what the relevant parts are and what their interrelations are. So it is easier to understand the operation of complex systems by considering simpler versions of them than by doing so directly. Furthermore, considering simple cases will suggest complexities that we can look for in sophisticated cases. I don't think there is any reason in principle why simple cases

should be so much easier to understand than complex cases - if we had more powerful brains, perhaps we would find complex cases nearly as perspicuous as we in fact find simple ones.

The third reason for considering simple mechanisms, like the first, is substantive. There are issues about what capacities simple mechanisms can have and how much complexity is required to have sophisticated capacities. In the case of many capacities - such as that for making subtle, precise discriminations - it is not obvious how a simple mechanism could have them, or even that it could have them. In such cases, showing how a simple mechanism can have the capacity shows that it is *possible* for such objects to have it, and by making the object with the capacity as simple as possible, we can see just how much complexity is needed for an object to have it. One of our ultimate goals is to show that we can apply the same distinctions that we ordinarily apply to the psychological states and actions of people to simple mechanisms. While, perhaps, we can apply the entire body of such distinctions only to people, if we can show that *each* distinction applies to *some* simple mechanism – so that we can *construct* concepts that apply (perhaps) only to people out of concepts each of which applies to some simple mechanism - we will have succeeded in giving a physicalistic reduction of human psychological states and action.

This concludes my introductory remarks. I have said what the problems are that the theory of indication I will develop aims to solve, and I have discussed the methodology I will follow. I turn now to the theory.

3.2 *Simple indication*

I will begin my discussion of indication by characterizing the intuitive notion, will next state a definition, and will illustrate the definition with a paradigm case. I will then go on to clarify and defend the theory. In section 3.2 I will be concerned primarily with cases of *simple* indication, cases in which an object indicates something as a result of being in a single state. In section 3.3 I will consider in detail cases of complex or *distributed* indication, cases in which an

object indicates something in virtue of combining multiple instances of simple indication in a way I will describe.

3.2.1 Definition

I will take the intuitive idea behind indication to be the following. As a result of a reasonably regular and reliable connection, a connection that holds when things are working as they normally do, the fact that a part of the world is in a certain state provides a basis for determining, carries the information, or *indicates* that the proposition that it is in a certain other state, or another part of the world is in a certain state, is true.⁹⁶ For instance, the fact that an alarm sounds may indicate that the air around it is smoky. For the alarm to indicate this, it must be true, when conditions are normal, that it would sound only if the air were smoky. Similarly, the fact that a thermometer reads 60 may indicate that the temperature is 60 degrees, the fact that a light is on may indicate that an appliance is on, and the fact that the light radiating to the earth from a star is shifted towards the red end of the electromagnetic spectrum may indicate that the star is moving away from the earth at a certain velocity. Thus we can think of the locution, the fact that Q indicates that P, as one kind of answer to the question, “how can you tell that P?”, an answer whose truth depends on the obtaining of a connection between a fact and a proposition.

I will now try to make precise the informal characterization of indication I have just given. As an initial attempt at a definition, I propose the following, based on an account of Robert Stalnaker.⁹⁷

- (I) The fact that an object or system *o* is in a state *a* indicates the proposition that *P* if and only if all of the following conditions are true:
 - (i) *o* has a set of relevant states *S* that form a partition, of which *a* is a member;

- (ii) the environment e of o has a set of relevant S' that form a partition;
- (iii) there is a one-one function f taking each element of S into a corresponding element of S' ;
- (iv) there is a set of normal conditions N ;
- (v) if every condition in N obtains, then, for every element s of S ,
 - (a) o would be in s only if e were in $f(s)$, and
 - (b) e 's being in $f(s)$ would cause o to be in s ;
- (vi) the proposition that e is in $f(a)$ implies P .

Note that as I understand the counterfactuals in (va) and (vb), and as I will understand counterfactuals more generally in this chapter, they can be true even if their antecedents are true. Also, I have suggested that the fact that an object is in one state can indicate that the same object is in a second state. For instance, the fact that a volume of gas in a closed container has a certain temperature can indicate that the gas is under a certain pressure. Accordingly, I will stipulate that the indicator itself can play the role of the environment in the definition.

To illustrate the definition, I will apply it to explain a paradigm case of indication. We suppose that the fact that an alarm sounds indicates that the air around it is smoky, and that what makes this true is that the following conditions hold:

- (i) The alarm has two relevant states, sounding and not sounding.
- (ii) The environment of the alarm - in this case, the air around it - also has two relevant states, smoky and not smoky.
- (iii) Each state of the alarm corresponds to a different state of the air: sounding to smoky, and not sounding to not smoky.

- (iv) There is a set of normal conditions. These are the possible facts that the alarm has a battery with sufficient charge, that it is in working order, that there is nothing else besides smoke that the air can contain that would set off the alarm, and so on.
- (v) If the normal conditions hold – that is, if the possible facts are actual - then:
 - (a0) the alarm would make noise only if the air were smoky,
 - (b0) the air's being smoky would cause the alarm to sound;
 - (a1) the alarm would not sound only if the air were not smoky,
 - (b1) the air's being not smoky would cause the alarm not to sound.
- (vi) The proposition that the air is smoky implies (trivially) that the air is smoky.

In the remainder of section 3.2, I will clarify the definition of indication I have given, and will consider some issues and problems that arise for it. This will set the stage for considering distributed indication in section 3.3.

3.2.2 States of indicator and environment

I want to consider first the states that clauses (i) and (ii) of our definition require the indicator and the environment to have. The clauses requires that both the indicator and the environment have a set of relevant states, and that both sets of states form partitions. That is, the states in both sets must be disjoint – neither the indicator nor the environment can be in more than one of its states at once; and the states must be exhaustive – the indicator must be in at least one of its states and the environment must be in at least one of its states. There will be many possible states of both the indicator and the environment that will *not* be relevant, so determining whether a putative case of indication satisfies our definition will involve distinguishing the relevant states from irrelevant ones.

An issue that arises about the relevant states of the indicator concerns the stipulation, in clause (i) of our definition of indication, that they be *intrinsic*. Must they be intrinsic, or can they be relational as well? Stalnaker, whom I have followed in defining indication, suggests they are intrinsic,⁹⁸ and it is easy to think of cases in which they are. Containing mercury that is at a certain level, having a certain number of rings, and sounding are indicating states that are intrinsic to, respectively, a thermometer, a cross-section of a tree, and an alarm. However, Israel and Perry suggest that the fact that a certain apple near the surface of the earth is left unsupported indicates that it will fall.⁹⁹ In this case, the indicating state, left unsupported, is not an intrinsic state of the apple but a relational state: whether the apple is left unsupported depends, not on facts just about the apple or its parts, but on facts about the relation of the apple to other things. Similarly, the fact that a fishing bobber is pulled beneath the surface of a pool of water may indicate that a fish has taken the bait, but this is a fact about the relation of the bobber to the surface of the water, and not just about the bobber or its parts.

The difference between the claim that indicating states are always intrinsic and the claim that they can also be relational is, I think, only a verbal difference. When we think of the indicating states of the bobber as relational, we are thinking of the bobber taken by itself as the indicating system. But, alternatively, we can think of the bobber as part of a larger indicating system that also includes the space above and below the bobber. If we think of the bobber as part of a larger indicating system, then the position of the bobber relative to the surface of the water is an intrinsic property of the larger system. Similarly, in the case of the apple, when we were thinking of the indicating state, left unsupported, as relational, we were thinking of the apple taken by itself as the indicating system. But in this case too we can alternatively think of the apple as part of a larger indicating system that also includes the region of space between the apple and the surface of the earth. Then the fact that the apple is left unsupported is a fact about an intrinsic feature of the larger system of which the apple is a part.

3.2.3 *The causal condition*

I have just discussed the states of the indicator and the environment mentioned in our definition of indication. In this section I will consider some issues that the causal condition in the definition raises, and will argue that it should be dropped from the definition.

Condition (vb1) in the smoke alarm example says that the air's not being smoky would cause the alarm not to sound. But saying that an *omission* by one event causes the *omission* of another sounds odd, and if one takes causation as a relation between events, then, since omissions are not events, they are not candidates for the relata. It is worth noting that not every case of indication requires assuming that omissions can be causes and effects. Suppose that a thermometer has two relevant states, red and blue, and that the fact that it is red indicates that the temperature is at least 70 degrees. This is so because, when conditions are normal, the thermometer would be red only if the temperature were at least 70 degrees, and the temperature's being at least 70 degrees would cause it to be red; and it would be blue only if the temperature were less than 70 degrees, and the temperature's being less than 70 degrees would cause it to be blue. Still, the alarm is intuitively a clear case of an indicator, and, as our definition now stands, indicators like it commit us to allowing that omissions can be causes and effects. One might argue that they can be, as, for instance, David Lewis has done.¹⁰⁰ But whether omissions can be causes or effects is still controversial, and our theory of indication would obviously be better off if it did not require a controversial assumption, even if we have an argument for it.

The problems of causation of and by omission that the causal condition raises are reasons to drop the causal condition from our definition of indication. But does dropping the causal condition have any costs? In particular, is the causal condition needed to rule out any cases that the counterfactual condition by itself does not rule out? The question we are asking is not simply whether there are any cases in which the counterfactual condition is satisfied and the causal

condition is not. To see what more we are looking for, suppose we find a case in which the counterfactual condition holds and the causal condition does not. Then one of the following three assertions about it will be true: (1) the intuitive notion of indication clearly *includes* the case; (2) the intuitive notion neither clearly *includes* nor clearly *excludes* the case; (3) the intuitive notion of indication clearly *excludes* the case. If (1) is true, then we *should* drop the causal condition from our definition since we do not want our definition to exclude a case that the intuitive notion of indication clearly covers. If (2) is true, then the case is irrelevant to our decision whether to drop the causal condition: whether the commonsense notion of indication we are trying to define includes or excludes the case should be seen as a consequence of our definition, and whether our definition includes or excludes the case has no bearing on its adequacy. It is only if (3) is true that we should not drop the causal condition, since we do not want our definition to include a case that the intuitive notion clearly excludes. So what we want to know is whether there are any cases in which the counterfactual condition holds and the causal condition does not, and, if so, whether those cases are ones in which it is intuitively clear that the indication relation fails to hold. Only such cases would be counterexamples to the contention that we can drop the causal condition from our definition.

Though I have no general argument that there are no counterexamples, I will consider two kinds of cases in which the counterfactual condition is satisfied and the causal condition is not. It will be clear, I think, that the intuitive notion of indication clearly covers those cases, and so that the causal condition should be dropped. In this way, I hope to put the burden of proof on one who claims that the definition requires the causal condition.

The first kind of potential counterexample I will consider to the contention that the causal condition can be dropped from our definition are cases in which a cause apparently indicates an effect. According to condition (vb) of our definition, the fact that an object *o* is in a state *a* indicates that the environment *e* is in a corresponding state *f(a)* only if *e*'s being in *f(a)* would

cause *c* to be in *a*. Thus, as our definition now stands, the event in virtue of which the indicated proposition is true must cause the event in virtue of which the indicating fact obtains. This is the case for the alarm: the fact that the alarm makes noise indicates that the air is smoky, and the air's being smoky causes the alarm to sound. The following example shows, however, that our ordinary notion of indication allows not only that an *effect* can indicate a *cause*, but also that a *cause* can indicate an *effect*. Suppose that a light switch is in one room and the light it controls in another, and that flipping the switch causes the light to go on. Then we might well determine whether the light is on by looking at the switch and seeing whether it is flipped, and it seems natural to say, if the switch is flipped, that the fact that it is indicates that the light is on. So the cause indicates the effect. Since in this kind of case the counterfactual condition is satisfied and the causal condition is not, and since it seems clear that the intuitive notion covers such cases, they provide evidence that the causal condition should be dropped.

The second kind of potential counterexample I will consider to the contention that the causal condition can be dropped from our definition of indication are cases in which two events that have a common cause apparently indicate each other via the cause. Suppose that flipping a switch causes both a machine to go on and a light to go on. Suppose, further, that the fact that the light is on indicates that the switch is flipped, and that the fact that the switch is flipped indicates that the light is on. Finally, suppose that from the two counterfactuals, the light would be on only if the switch were flipped, and the switch would be flipped only if the machine were on, we can infer that the light would be on only if the machine were on. Then we can use the fact that the light is on to tell that the machine is on, and it seems natural to say that the fact that the light is on indicates that the machine is on. However, though the light's being on and the machine's being on are effects of a common cause, the light's being on is neither a cause nor an effect of the machine's being on. So though this example satisfies the counterfactual condition and fails to

satisfy the causal condition, it is also intuitively a clear case of indication, and so provides further motivation for dropping the causal condition.

I have considered two potential counterexamples to the contention that we can drop the causal condition from our definition of indication, but have argued that both cases show, instead, that we *should* drop the causal condition. An alternative to dropping the causal condition would be to give a disjunctive definition of indication – one that allows that a cause can indicate an effect, and that effects of a common cause can indicate one another, in addition to allowing that an effect can indicate a cause. But a disjunctive definition would be an ad hoc response that would result in a much more complicated definition, and would still require making the controversial assumption that omissions can be causes and effects. So I will accommodate the cases we have just considered by modifying clause (v) of our definition as follows:

- (v') if the normal conditions obtain, then, for every element s of the set of relevant states of the indicator o , o would be in s only if the environment were in the corresponding state $f(s)$.

A theory of indication with no causal requirement is consistent with our assumption that the core idea underlying the intuitive notion of indication is that one can tell what is going on in one part of the world by observing a more salient or accessible part of the world, without regard to why one can tell. But if this is the central idea underlying indication, why did it seem appropriate to include the causal condition in our definition? The reason, I think, is that in many cases of indication the satisfaction of the causal condition explains *why* the counterfactual relation holds. In those cases the causal condition is a normal condition. But we should not confuse a condition that is a common normal condition with a condition that belongs in the definition.

3.2.4 The logical condition

Since we have dropped the causal condition from our definition of indication, the only kind of relation required for a fact to indicate a proposition is a logical one. I will first clarify the nature of the logical relation, and will then discuss some consequences of the condition in the definition that requires that the logical condition hold.

According to clause (v'), if the fact that an indicator is in a relevant state a indicates that the environment is in a corresponding relevant state $f(a)$, then not only must it be true that the indicator would be in a only if the environment were in $f(a)$. It must also be true, for every other relevant state s of the indicator, that the indicator would be in s only if the environment were in the corresponding relevant state $f(s)$. That is why the "only if" in clause (va) must be interpreted counterfactually: to know whether a putative case of indication satisfies the definition, one must know not only whether the conditional in (va) holds in the case of the actual relevant state of the indicator, but also whether it holds in the case of the other relevant states of the indicator, states the indicator can have but, since the relevant states of the indicator are disjoint, does not actually have. Because the "only if" is interpreted counterfactually, I assume that to say that P only if Q is not to say, in terms of "if," that Q if P ($P > Q$, where ">" stands for the counterfactual connective). Rather, it is to say that it is false that P if not- Q ($\sim(\sim Q > P)$), and this I assume to be equivalent to not- P if not- Q ($\sim Q > \sim P$).

I turn now to some consequences of the logical condition. Recall our earlier example, in which the fact that an alarm sounds indicates that the air is smoky. In that example, if conditions are normal, not only must it be true that the alarm would sound only if the air were smoky; it must also be true that the alarm would *not* sound only if the air were *not* smoky. Thus it is clear that the example satisfies clause (v') of our definition: the relevant states of the alarm are sounding and not sounding, and, if conditions are normal, then, for every relevant state s of the alarm, the alarm would be in s only if the environment were in the corresponding state.

The requirement that these counterfactuals must hold between each member of the set consisting of the propositions that the indicator is in each of its relevant states, and a corresponding member of the set consisting of the propositions that the environment is in each of its relevant states, has two consequences I want to note. We can characterize the first as follows. Suppose that the fact that an indicator is in a state a indicates that the environment is in a corresponding state $f(a)$. Then there exists a set of relevant states of the indicator $S = \{s_0, s_1, \dots\}$ of which a is a member, and a set of corresponding relevant states of the environment $S' = \{f(s_0), f(s_1), \dots\}$ of which $f(a)$ is a member. We can define a set of possible facts $F = \{f_0, f_1, \dots\}$, where each f_i is the possible fact that the indicator is in s_i . We can also define a set of propositions $P = \{p_0, p_1, \dots\}$, where each p_i is the proposition that the environment is in $f(s_i)$. Since we have supposed that some element of the set of possible facts F indicates the corresponding member of the set of propositions P , it follows, by clause (v'), that for every possible fact f_i in F , the proposition that f_i obtains would be true only if the corresponding proposition p_i were true. And it follows from this, by our definition of indication, that for every possible fact f_i , if it were actual, then it would indicate p_i .

This is the first consequence of clause (v') I want to note: if the fact that the indicator is in one of its relevant states indicates that the environment is in the corresponding state, then if the indicator were in any of its other relevant states, the fact that it was would indicate that the environment was in the corresponding state. In the alarm example, since the fact that the alarm sounds indicates that the air is smoky, we can infer that if the alarm did not sound, then the fact that it did not would indicate that the air was not smoky. We might say, then, that there is indication only in the context of an indicating *system*, a system capable of indicating multiple, mutually exhaustive and exclusive propositions about the environment, corresponding to the mutually exhaustive and exclusive relevant states of the environment. (Note that though the environment has only two relevant states in the example, it could have more.)

The second consequence I want to note of the requirement in clause (v') that, for every relevant state s of the indicator, the indicator would be in s only if the environment were in the corresponding relevant state, is this: if the fact that an indicator is in a relevant state a indicates that the environment is in the corresponding relevant state $f(a)$, then whether the indicator is in a *depends counterfactually* on whether the environment is in $f(a)$. That is, both of the following counterfactuals will be true: if the environment were in $f(a)$ then the indicator would be in a , and if the environment were not in $f(a)$ then the indicator would not be in a .

We can show this as follows. Since the relevant states of the indicator and the environment form partitions, if the fact that an indicator is in one of its relevant states indicates that the environment is in the corresponding one of its relevant states, then our definition implies not only that every member of the set consisting of the propositions that the indicator is in each of its relevant states would be true *only if* the environment were in the corresponding relevant state, but also that every member of the set of propositions about the indicator would be true *if* the corresponding proposition about the environment were true.¹⁰¹ So, if we suppose that the fact that the indicator is in a indicates that the environment is in $f(a)$, then it will be true not only that the indicator would be in a *only if* the environment were in $f(a)$, but also that the indicator would be in a *if* the environment were in $f(a)$. Since we are assuming that P only if Q is equivalent to not- Q if not- P , it follows from the first counterfactual that the indicator would not be in a if the environment were not in $f(a)$. So whether the indicator is in a depends counterfactually on whether the environment is in $f(a)$. For instance, since the fact that the alarm sounds indicates that the air is smoky, both the following counterfactuals are true: if the air were smoky then the alarm would sound, and if the air were not smoky then the alarm would not sound. So now that the causal condition is gone, our definition explains indication in terms of counterfactual dependence.

There are, however, cases of indication in which it may seem that clause (v') is not satisfied. That is, in these cases, it may seem that there is some relevant state s of an indicator,

and a corresponding relevant state of the environment $f(s)$, such that it is false that the indicator would be in s only if the environment were in $f(s)$. Thus these cases seem to show that there can be indication outside the context of an indicating system, and without causal dependence.

Consider the following example. As before, an alarm has two relevant states, sounding and not sounding, and the air has two corresponding states, smoky and not smoky. There is a set of normal conditions, but in this example we exclude the condition that an electrical circuit is closed, even though the air's being smoky would not cause the alarm to sound unless the circuit were closed. We suppose that the normal conditions obtain, and the alarm sounds. Then it is clear that the fact that the alarm sounds indicates that the air is smoky, and that this counterfactual is true: the alarm would sound only if the air were smoky. But it is also clear that this counterfactual is false: the alarm would not sound only if the air were not smoky. It is false since even if the air were smoky, the alarm would not sound if the circuit were not closed. It follows that the possible fact that the alarm does not sound, if it were actual, would not indicate that the air was not smoky. And though the counterfactual, if the air were not smoky then the alarm would not sound, is true, the counterfactual, if the air were smoky then the alarm would sound, is false. So whether the alarm sounds is not dependent counterfactually on whether the environment is smoky. One might conclude that this is a case of indication that our definition cannot explain, a case in which there is indication without an indicating system, and without counterfactual dependence.

It would, however, be a mistake to reach this conclusion. What led to it was identifying the wrong set of states of the environment as the *relevant* states of the environment. And what led to the misidentification was considering the states corresponding, not to the *strongest* proposition indicated in the situation, the indicated proposition that implies all the other indicated propositions, but to a weaker proposition, a proposition that the strongest proposition implies and that does not imply the strongest proposition. That a single fact will indicate many different propositions is a consequence of our definition: by clause (vi), if a fact indicates a proposition,

then it also indicates every proposition that proposition implies. (So though I have been speaking of *the* proposition that a fact indicates, it would be more accurate to speak of the *strongest* proposition that a fact indicates, or of the *set* of propositions that a fact indicates.) In the example I have described, the strongest proposition indicated is the proposition that the air is smoky *and* the circuit is closed. The set of states corresponding to this proposition – the *relevant* states of the environment – are the state that it asserts to obtain, smoky and the circuit is closed, and the mutually exclusive and exhaustive state, not smoky or the circuit is not closed. Both counterfactuals associated with the relevant states are true: the alarm would sound only if the air were smoky and the circuit were closed, and the alarm would not sound only if the air were not smoky or the circuit not closed. As a result, not only does the fact that the alarm sounds indicate that the air is smoky and the circuit closed, but if the possible fact that the alarm does not sound were actual, then it would indicate that the environment was in its other relevant state, not smoky or the circuit is not closed. So there is an indicating system. Further, both these counterfactuals are true: if the air were smoky and the circuit closed, then the alarm would sound; and if the air were not smoky or the circuit not closed, then the alarm would not sound. So whether the alarm sounds depends counterfactually on whether the air is smoky and the circuit is closed.

Thus our definition allows for two different sorts of relations between a fact and a proposition that it indicates. The proposition may be the strongest proposition that the fact indicates, or it may be one of many weaker propositions. We will say, in the case of the strongest proposition, that the fact *strongly* indicates it, and, in the case of a weaker proposition, that the fact *weakly* indicates it.¹⁰² Thus the fact that the alarm sounds strongly indicates that the air is smoky and the circuit is closed, and it weakly indicates that the air is smoky. The main difference, from the user's point of view, between strong and weak indication is that weak indication, but not strong indication, allows false negatives. In normal conditions, the alarm *cannot* fail to sound when there is smoke and the circuit is closed, but the alarm *can* fail to sound when there is

smoke. Other things being equal, we prefer strong indication to weak indication, but, because of resource constraints, other things are not always equal. It may be easier, cheaper or faster to rely on weak indication, and expedience may encourage a strategy of first relying on weak indication, and then, if the result is negative, confirming the result by relying on strong indication. For instance, if in our example the circuit is mostly closed but sometimes open, one might determine whether there is smoke by relying for the most part on the alarm. But during long periods when the alarm does not sound, one might, to be sure there is no smoke, inspect another indicator from time to time, an indicator that is harder to use but, if there is smoke, strongly indicates that there is.

I will conclude my discussion of the logical condition in our definition of indication by considering the consequences for the symmetry of indication of analyzing it in terms of counterfactual dependence. Since indication as we construe it is a relation between a fact and a proposition, it relates different kinds of things and cannot be symmetric. However, we can define indication in terms of a relation between propositions – call it *propositional indication* or *p-indication* – that has a form that allows it to be symmetric. We do so as follows. The fact that an indicator o is in a relevant state a indicates that the environment e is in the corresponding relevant state $f(a)$ if and only if it is a fact that o is in a , and the proposition that o is in a p -indicates that e is in $f(a)$. Now is p -indication symmetric? The answer is no, since counterfactual dependence is not symmetric. Suppose that the fact that I act in a certain way indicates that I had certain prior beliefs. This implies, by clause (v'), that whether I act in that way depends (counterfactually) on whether I had those beliefs. But from this it does *not* follow that whether I had those beliefs depends (counterfactually) on whether I act in that way; for one thing, it does not seem right to say that if I acted in a different way, than I would have had different prior beliefs. This shows that one cannot infer, from the proposition that I act in a certain way p -indicates that I had certain prior beliefs, that the proposition that I had those prior beliefs p -indicates that I act in that way,

which seems intuitively to be the right result. So we cannot conclude that the fact that I had those beliefs indicates that I act in that way.

3.2.5 *The xerox principle*

I have just considered some consequences of the logical condition in our definition of indication. I want to consider now what the consequences of the logical condition are for the truth of the *xerox principle*, which says: “if A carries the information that B, and B carries the information that C, then A carries the information that C.”¹⁰³ According to Fred Dretske, the xerox principle is a fundamental principle of information that any theory of information should preserve. It is clear why one might think so: we commonly rely on chains of indicators to obtain information about remote or inaccessible times or places, as when we read the newspaper or listen to the radio, or make use of other *information flow systems*,¹⁰⁴ and it is the validity of the xerox principle that warrants doing so. The xerox principle is, of course, a version of the claim that indication is transitive. I will consider what the consequences of our definition of indication are for the transitivity of indication, and whether they are consistent with the commonsense notion we are trying to define.

Since we construe indication as a relation between a fact and a proposition, it can no more be transitive than it can be symmetric. So we will consider whether p-indication, rather than indication proper, is transitive. As our definition of indication now stands, we have analyzed it – and so p-indication – in terms of one other relation, the counterfactual relation. So whether p-indication is transitive thus depends on whether the counterfactual relation is transitive. But the counterfactual relation is not always transitive, so our definition of indication has the consequence that p-indication is not always transitive. This consequence of our definition seems consistent with the intuitive notion of indication. I suggested earlier that the main idea underlying the intuitive notion is that, as a result of a reasonably reliable connection between a fact and the

truth of a proposition, the fact provides a basis for concluding that the proposition is true. But if it is false that the fact would obtain only if the indicated proposition were true, then we have no basis for concluding that the indicated proposition is true even if we know that the fact obtains.

The counterfactual relation is nevertheless transitive in many cases. For while the counterfactual relation is not invariably transitive, the following closely related inference schema is valid.¹⁰⁵ Suppose that B if A, and that C if both A and B. Then it does follow that C if A. In terms of “only if” rather than “if”: from B only if C, and A only if either B or C, we can conclude A only if C. So though it is not true that p-indication is invariably transitive, it is true that if (i) A p-indicates B and B p-indicates C, and if (ii) the premisses of the closely related inference are true, then A p-indicates C. If, in addition, (iii) the possible fact corresponding to A is actual, then it indicates C. Suppose, for instance, that the fact that a loudspeaker emits a certain sound indicates that a microphone picked up the sound, and the fact that the microphone picked it up indicates that someone said “nevermore.” Then (i) the proposition that the speaker emits the sound p-indicates that the microphone picked it up, and the proposition that the microphone picked it up p-indicates that someone said “nevermore”; and (iii) it is a fact that the speaker emits the sound. Further, in this case (ii) the microphone would have picked up the sound only if someone had said “nevermore,” and the speaker would emit the sound only if either the microphone had picked it up or someone had said “nevermore.” So the fact that the speaker emits the sound indicates that someone said “nevermore.”

Since indication as we have defined it is not invariably transitive, it is at odds with the xerox principle if we take the principle to be invariably true. But even if the xerox principle is not always true on our definition, it is also not always false, and holds in many cases. In this connection it is important to recognize that while many indicators occur naturally and hold no special interest for people, others are appropriated for use by people, or are human artifacts. When we select or build an indicator, we can ensure that transitivity holds, as we do in the case of

information flow systems. This suggests that we should view the xerox principle not as a principle that governs every case in which information is transmitted or something is indicated, but as a normative principle that governs a well-behaved subset of those cases.

This completes my discussion of the xerox principle. In the next section I will discuss the normal conditions required in clauses (iv) and (v) of the definition, the conditions under which the counterfactuals required for indication are true.

3.2.6 Normal conditions

I will defend two claims about the normal conditions: first, that the set of normal conditions determines the set of propositions that a fact indicates, and second, that the normal conditions are the prevailing or statistically common conditions in the environment in which the fact indicates the set of propositions. Together, these two conditions imply that the statistically common conditions in an environment determine the set of propositions that a fact that obtains in the environment indicates. From this it follows that we can give a reductive physicalistic explanation of what determines the propositions.

I will first argue that when a fact indicates a set of propositions, what determines the members of the set of propositions is the associated set of normal conditions. Consider a doorbell. The fact that a doorbell rings indicates that a person has pressed it. The reason is that in the actual world only people tend to press doorbells, so, in normal conditions, a doorbell would ring only if a person pressed it. But suppose that the normal conditions were different, and that only dogs tended to press doorbells. Then, in those normal conditions, a doorbell would ring only if a dog pressed it, and if a doorbell rang, the fact that it did would indicate that a dog had pressed it (provided the other conditions required for indication were satisfied). Since both the proposition that a person has pressed a doorbell and the proposition that a dog has pressed a doorbell imply that a doorbell has been pressed, the latter is a proposition that the fact that a doorbell rings would

indicate relative to both sets of normal conditions, since, by clause (vi) of our definition, a fact indicates every proposition implied by any proposition it indicates. But though the proposition that the doorbell has been pressed is more general than either indicated proposition that implies it, and is indicated relative to a wider range of sets of normal conditions than either of them, it is clear that it too is indicated by the fact that a doorbell rings only relative to certain sets of normal conditions. For suppose that doorbells operated differently from how they actually do, so that they were made to ring, not by pressing them, but by touching them. Then if a doorbell rang, the fact that it did would indicate, not that it had been pressed, but that it had been touched.

As the examples suggest, to some extent one can change the propositions that a fact indicates by changing the normal conditions. For instance, one might put a doorbell in a location accessible only to dogs and train dogs to use it. Then the fact that it rang would indicate that a dog had pressed it. But this does not mean that one can create normal conditions that would make a fact indicate any proposition one wants. One cannot create normal conditions in which, if an ordinary doorbell rang, that fact would indicate that it had *not* been pressed, since such normal conditions would require violating causal laws. Nevertheless, it may be true that in a possible world with different causal laws from ours the fact that a doorbell rang would indicate that it had *not* been pressed.

I have just argued that the propositions that a fact indicates are relative to specific sets of normal conditions, so that the same fact can indicate two inconsistent propositions given two inconsistent sets of normal conditions. But what determines what the normal conditions are? One possible answer is to say that the normal conditions are any conditions a theorist chooses (or, perhaps, the normal conditions are the relevant laws of nature plus any further conditions consistent with the laws that the theorist chooses). But this answer is unacceptable for a physicalistic reduction of indication, since it would lead to a definition of indication (partly) in

terms of what a person chooses, rather than a definition in terms of natural facts about indicators and their environments.

To see how we can define normal conditions in terms of natural facts, consider an example I gave earlier: the fact that an apple is left unsupported near the surface of the earth indicates that it will fall. The apple has a set of relevant states, left supported and not left unsupported, and it also has a set of corresponding relevant states, will be falling and will not be falling. Further, if the fact that an apple is left unsupported indicates that it will fall, then it follows from our definition of indication that, in normal conditions, the following counterfactual holds: the apple would be left unsupported only if it would fall. On our definition of indication, the normal conditions are those conditions such that, if they obtain, then the counterfactual holds. So, in this case, the normal conditions are that the earth is surrounded by a gravitational field, and that the apple has mass. It is obvious that these conditions are the statistically common or prevailing conditions – it is a general fact about apples that they have mass, and it is a universal law that massive bodies like the earth generate gravitational fields.

If we consider the most straightforward way to come up with an indicator for a certain proposition in a given environment – and so to tell whether, in that environment, the fact corresponding to the proposition obtains – it should come as no surprise that the normal conditions are the prevailing conditions. The most straightforward way to tell whether a fact obtains is obviously to exploit a correlation that already exists in the environment between that fact and another fact the agent can more easily determine to obtain. It is clear that an indicator like this can be reliable, and useful, only if the corresponding normal conditions are the prevailing conditions, and it is also clear that many indicators are like this. In these cases it seems reasonable to identify the normal conditions with the prevailing conditions.

However, an agent might need to tell whether a fact obtains in the environment that has no existing correlation with any other fact that can obtain in the environment whose obtaining she

can recognize. In the case of a fact with no existing correlation she can exploit, the prevailing conditions in the environment do not provide a set of normal conditions in virtue of which any other fact in the environment indicates the proposition that the fact she is interested in obtains. I argued earlier that what a fact indicates is determined by the set of normal conditions associated with it, and that (up to a point) an agent can change what a fact indicates by changing the normal conditions. I will now give an example to show that an agent can create an indicator to tell whether a fact obtains that has no existing correlation she can exploit by changing the prevailing conditions in the environment. If I am right that the example shows this, then since an agent can change what a fact indicates by changing the prevailing conditions just as (I argued earlier) she can change what it indicates by changing the normal conditions, the example provides further evidence that the normal conditions should be identified with the prevailing conditions.

Suppose that an agent needs to identify flying insects in an environment in which the only things that can fly are insects, but there is no existing correlation she can exploit between the fact that an object is a flying insect and any other fact she can determine to obtain. In the environment we are considering, if the prevailing conditions obtain, then an object would land on a surface suspended above the ground only if the object were either something that had fallen from the sky or a flying insect. But if the agent places a canopy over the surface, she will create a “microenvironment” in which the prevailing conditions deviate from the prevailing conditions in the environment at large, and a different counterfactual will hold if those prevailing conditions obtain: if the prevailing conditions under the canopy obtain, an object would land on a surface suspended above the ground only if it were a flying insect and *not* something that had fallen from the sky. So the fact that an object lands on a surface off the ground indicates one thing in the environment at large, and another thing in the microenvironment under the canopy, and we have explained why in terms of a difference between the prevailing conditions in the environment at large and those in the microenvironment.

3.3 Distributed indication

The example I just gave is an example of *distributed indication*. In distributed indication, an indicator is made up of other indicators in such a way that the strongest proposition it indicates is a function of the strongest propositions its constituents indicate. As a result, the strongest proposition the distributed indicator indicates is stronger or more specific than the strongest propositions its constituents indicate. In our last example, the fact that an object lands on a surface suspended off the ground indicates that it has fallen from the sky or is a flying insect, and the fact that an object is beneath a canopy indicates that it has not fallen from the sky. But the fact that an object is on a surface suspended off the ground *and* is beneath a canopy indicates that it is a flying insect *and* has not fallen from the sky. As the example suggests, distributed indicators can determine that objects are of certain specific types. Once a distributed indicator has determined that an object is of one specific type, it may go on to indicate an even more specific or precise proposition. For instance, once an indicator has identified an object as a flying insect, it may identify the object as a particular kind of flying insect, and once it has done that, it may identify a part of a surface as the area between the eyes of the insect. More generally, an agent can combine indicators to form distributed indicators capable of distinguishing objects of one type from objects of all other types in a given environment, and so capable of indicating propositions of an arbitrary degree of specificity.

Since distributed indicators can indicate very specific propositions, and since they do so by combining other, simpler indicators that indicate less specific propositions, they are especially important for the project of giving a physicalistic reduction of indication, and I will consider them in detail. In section 3.3.1 I will discuss at length an example of an orchid flower that uses distributed indication to distinguish bees from other objects in its environment. The flower

illustrates just one kind of structure that a distributed indicator can have, to which there is an important alternative. I will describe and contrast the structures in section 3.3.2.

3.3.1 Distributed indication in an orchid

In this section, I will describe the mechanism of orchid fertilization in general. This will set the stage for describing how fertilization occurs in the specific variety of orchid I will consider. I will then apply our definition of indication to explain what the fact that one part of a flower of that kind of orchid is in one local state indicates, and what the fact that another part of the flower is in another local state indicates. Finally, I will apply our definition to explain what is indicated by the fact that the flower is in the global or distributed state that the local states form. First, however, I want to say how the example of distributed indication I will discuss bears on the project of giving a physicalistic reduction of indication in general, and of indication in people in particular.

The example I will discuss bears on the larger project in two ways. First, the example will show how a simple mechanism can indicate a relatively specific or precise proposition, and so how an object can do so just by having parts that have certain shapes and sizes, move in certain ways, stand in certain spatial relations to one another, and have certain causal powers as a result of those features. Thus the example will show that it is at least *possible* to have an internal representation that can figure in a fine-grained, sophisticated belief, a belief of a degree of subtlety and precision that we ordinarily associate only with people, without having any capacities, such as a capacity for language, that we do not ordinarily associate with simple mechanisms. So one thing I hope to accomplish with the example is to present a simple alternative to the view that internal or mental representations must have linguistic form, and so to go some distance towards demystifying such representations. The second way in which the example I will discuss bears on the larger project is that it will show how to give a physicalistic

reduction of a relatively sophisticated case of indication. So the example will suggest what a physicalistic reduction of sophisticated cases of indication might look like, as well as take a step towards giving a physicalistic reduction of indication in general.

I turn now to the general mechanisms of orchid fertilization. Many orchids insure that only creatures of one particular type, creatures that tend to have the same size and shape, can fertilize them. They do so by allowing only creatures of the target type to reach their pollen, which in orchid flowers is typically formed into a single, sticky mass. Pollination occurs when a pollinator makes its way to one flower, where a pollinium gets stuck to it, and then makes its way to another flower, where the pollinium is brushed off onto the stigma and germinates. By allowing only creatures of the same general size and shape to reach its pollinium, an orchid flower can insure (in its native habitat) that it will be attached only to a specific part of the creatures. The precise placement of the pollinium helps guarantee that it will be brushed off only onto the stigma of a flower of the same variety, and not onto anything else.

The particular flower I will describe does not actually exist, but is a composite of ones that do. So though no variety of orchid has flowers with all the indicating mechanisms I will describe, every mechanism is one that flowers of some variety of orchid actually have.¹⁰⁶ The flower we are considering guarantees (in normal conditions) that only bees can depress the *pollinium lever*. When the lever is depressed, the *pollinium stalk* springs loose, and its trajectory carries the pollinium, located at the tip of the stalk, to the part of the bee's head between its eyes, where the pollinium sticks. The pollinium lever is located inside the *attachment holding chamber*, itself located inside the flower. When the bee presses the lever and the stalk springs loose, the bee, as a result of being lured inside the chamber by a trail of colored dots, is properly positioned inside it so that the pollinium will strike the bee between the eyes: as a result of following the dots, the bee reaches a point where its front legs depress the pollinium lever with the bee facing in the right direction. What guarantees that only bees will follow the colored dots into the chamber,

and not birds – which we suppose the dots also attract – is that birds are too large to fit inside the opening to the chamber. Such devices as paths of colored dots, and openings of certain sizes, can be multiplied indefinitely as needed to discriminate among objects normally found in the orchid's habitat. (As a result, actual orchids are capable of distinguishing not only bees from other creatures, but bees of one species from bees of other species, and even males of the species from females.) Thus it is guaranteed, if conditions are normal, that an object would depress the pollinium lever only if it were a bee since only bees have the right combination of traits to reach the chamber and depress the lever – only they are small enough to enter the chamber, are heavy enough to depress the lever, are attracted inside the chamber by the colored dots, are attracted to the flower by its smell, inhabit the same ecosystem as the orchid, and so on.

As I have described the example, if an orchid's pollinium lever is depressed by an object, then the fact that it is will indicate that the object is a bee. The fact will indicate the proposition since the lever is in a specific microenvironment and stands in specific spatial relations to other parts of the orchid. If the lever were not hidden away inside the flower, and if the other constituent indicators did not guarantee (in normal conditions) that the only objects capable of reaching and depressing the lever were bees, then the fact that the lever is depressed by an object would indicate only a more general proposition – that some object in the orchid's normal environment heavy enough to depress the lever had done so. Thus the lever is an element of a distributed indicator, and what the fact that the *flower* is in a certain *global* state indicates is a function of what the facts that various *parts* of the flower are in certain *local* states indicate. I will show first how to apply our definition of indication to explain what the fact that one part of the flower is in one local state indicates, and what the fact that another part of the flower is in another local state indicates, and then how to apply it to explain what the fact that the flower is in the global state that is a function of the two local states indicates.

The constituent indicators of the flower that I will consider are the pollinium lever and the opening to the holding chamber. I will now apply the definition of indication to explain what each constituent indicates in isolation from the other. Thus I will consider what the fact that the lever was depressed by an object would indicate if the pollinium lever were by itself outside the flower, though in the same environment. And, similarly, I will consider what the fact that the opening of the flower was entered by an object would indicate if the opening occurred by itself in isolation from the rest of the flower, though in the same environment. We suppose that the fact that the pollinium lever is depressed by an object indicates (or, rather, would indicate if the pollinium lever occurred outside the flower) that the object is a bird or a bee.

We explain this as follows, letting “D,” “I,” “B,” “G,” and “N” refer to the respective states: depressed by an object, bird, bee, gnat, and no object:

- (i) The pollinium lever has two relevant states: D, not-D.
- (ii) The object that depresses, or does not depress, the lever has two relevant (disjunctive) states: I or B, G or N.
- (iii) There is a one-one function from each state of the lever to each state of the object: D corresponds to I or B, not-D corresponds to G or N.
- (iv) There is a set of normal conditions. I will not try to specify all of them, but they include the conditions that the pollinium lever is not a mutant that requires an abnormally strong force to depress it or that can be depressed by an abnormally weak force, and that no gnats in the lever’s environment are abnormally heavy and no birds or bees are abnormally light.
- (v) If the normal conditions obtain, then the lever would be in D only if the object were in I or B, and the lever would be in not-D only if the object were in G or N.

- (vi) The proposition that the object is in I or B implies that the object is a bird or a bee.

Note that for the fact that the pollinium lever is depressed to indicate that the environment contains a bird or a bee, we must make two further assumptions regarding the normal conditions. First, in keeping with our strategy of considering what the two constituent indicators would indicate if they occurred outside the flower, we ignore the fact that the size of the opening to the flower restricts the kinds of objects that can reach and step on the pollinium lever, and so we exclude from the normal conditions the fact that the opening to the flower filters out certain types of objects. Thus, for purposes of our analysis, we make an exception here to the principle that the normal conditions are the statistically common conditions. Second, we assume for simplicity that the only kinds of objects that normally occur in the flower's environment that are heavy enough to depress the lever are birds and bees, and that the only objects that are light enough to fail to depress the lever when positioned on it are gnats. This assumption we include in the set of normal conditions.

Next we consider what the fact that the opening to the flower is entered by an object indicates. It indicates, we suppose, that the object is a gnat or a bee, and we explain this as follows, using the same abbreviations as before, and letting "E" stand for the state, entered by an object:

- (i) The opening to the orchid flower has two relevant states: E, not-E.
- (ii) The object that enters, or does not enter, the opening has two relevant states: G or B, I or N.
- (iii) Each state of the opening corresponds to a different state of the object: E to G or B, and not-E to I or N.

- (iv) There is a set of normal conditions. These are roughly parallel to the normal conditions in virtue of which the fact that the lever is depressed by an object indicates that the object is a bird or a bee.
- (v) If the normal conditions obtain, then the opening would be in E only if the object were in G or B, and the opening would be in not-E only if the object were in I or N.
- (vi) The proposition that the object is in G or B implies that the object is a gnat or a bee.

We have applied our definition of indication to explain what the fact that the lever is depressed by an object indicates about the object, and what the fact that the opening is entered by an object indicates about the object, where we considered each indicator in isolation from the other, and from the rest of the flower. Now we apply the definition to explain what the two indicating facts indicate together as components of a distributed indicator. When we considered each constituent indicator in isolation, we considered what *single* property the fact that each constituent is in a certain state indicates that the object has – what single property the fact that the lever is depressed by an object indicates that the object has, and what single property the fact that the opening is entered by an object indicates that the object has. Now we will consider what *conjunction* of properties the fact that the lever is depressed by an object and the fact that the opening is entered by the same object indicates that object has.

We start by considering the relevant states of the indicator, which in this case is the distributed indicator consisting of the two constituent indicators, the lever and the opening. Since each constituent has two relevant states, the distributed indicator has four relevant states corresponding to the four ways in which the two local indicators can be combined. So

- (i) The distributed indicator has a set of relevant states: E and D, not-E and D, E and not-D, not-E and not-D.

Note that given the way the orchid is constructed, not all of these complex states are ones that can be realized if conditions are normal. This, however, does not matter to the analysis.

Next we consider the relevant states of the object that fits, or does not fit, inside the opening and that depresses, or does not depress, the pollinium lever. As in the case of the relevant states of the distributed indicator itself, the environment of the distributed indicator has four relevant states corresponding to the four ways in which the two relevant states of the environment of each local indicator can be conjoined:

- (ii) G or B, and I or B; I or N, and I or B; G or B, and G or N; I or N, and G or N.

Simplifying the four relevant states of the environment of the distributed indicator,¹⁰⁷ we can rewrite (ii) as:

- (ii') B, I, G, N.

The rest of the conditions in this case of indication are as follows.

- (iii) Each state of the distributed indicator corresponds to a different state of the object: E and D to B, not-E and D to I, E and not-D to G, not-E and not-D to N.
- (iv) There is a set of normal conditions. This is the union of the sets of normal conditions corresponding to each constituent indicator, plus the condition that the

object that passes through, or does not pass through, the opening is identical to the object that depresses, or does not depress, the pollinium lever.

- (v) If the normal conditions obtain, then the distributed indicator would be in E (at one time) and D (at another time) only if the object were in B, the indicator would be in not-E and D only if the object were in I, the indicator would be in E and not-D only if the object were in G, and the indicator would be in not-E and not-D only if the object were in N.
- (vii) The proposition that an object is in B implies that the object is a bee.

3.3.2 Sequential and parallel distributed indication

In the case of the orchid flower, the constituent indicators operate sequentially. But distributed indicators can have constituent indicators that operate in parallel. To bring out the difference between sequential distributed indicators and parallel distributed indicators, I will consider an example of a rattlesnake, based on an example of Patricia and Paul Churchland.¹⁰⁸ A kind of rattlesnake lives in an environment in which there are many moving things that are not prey, and many warm things that are not prey, but very few warm moving things that are not prey, so the snake can identify prey by identifying things that are both warm and moving. We can distinguish two different distributed indicators the snake might use to identify things that are both warm and moving, a sequential distributed indicator and a parallel distributed indicator. I will describe the structure of each distributed indicator, and will then compare the structures.

The sequential distributed indicator works as follows. Suppose the snake has a neuron sensitive to movement – call it an *m-neuron* – which receives input from the eyes, and a neuron sensitive to heat – call it an *h-neuron* – which receives input from the pits and from the m-neuron. Both neurons have the same two relevant states, firing and not firing. However, in the resting state of the sequential distributed indicator made up of the m-neuron and the h-neuron, the h-

neuron is inhibited and cannot fire, regardless of the input it receives from the pits. The m-neuron, on the other hand, would fire (in normal conditions) if and only if the eyes perceived a moving thing. The firing of the m-neuron disinhibits the h-neuron, and, if disinhibited, the h-neuron would fire (in normal conditions) if and only if the eyes perceived a moving thing. But since the h-neuron can fire in response to a warm thing only when the warm thing is moving, the fact that the h-neuron fires indicates that there is prey.

Like the sequential distributed indicator, the parallel distributed indicator is made up of an h-neuron and an m-neuron, the neurons have the same two relevant states, and the m-neuron receives input from the eyes. But in this distributed indicator, the h-neuron receives input only from the pits and not from the m-neuron, and in the resting state the h-neuron is not inhibited from firing. The m-neuron would fire (in normal conditions) if and only if there were a moving thing, the h-neuron would fire (in normal conditions) if and only if there were a warm thing, and both neurons would fire (in normal conditions) if and only if there were a thing that was both warm and moving. So the fact that both neurons fire indicates that there is prey.

The key difference between the sequential distributed indicator and the parallel one is the following. In the sequential indicator, the indicating state that corresponds to the presence of prey – having the h-neuron fire – obtains if and only if a single *local* state, a state of a constituent of the distributed indicator, is instantiated. But in the parallel indicator, the indicating state that corresponds to the presence of prey – having *both* the m-neuron *and* the h-neuron fire – obtains if and only if a *global* state that consists of the local states of both constituent indicators is instantiated. This difference becomes relevant when we consider what is involved in *using* the two distributed indicators. Suppose the snake has a neuron whose firing causes the snake to strike – call it an *s-neuron* – and that it fires in response to appropriate input from a distributed indicator. If the input is from the parallel indicator, the s-neuron must determine the states of both

the m-neuron and the h-neuron. But if the input is from the sequential indicator, the s-neuron need determine only the state of the h-neuron.

I can see no general reason to prefer one kind of distributed indicator to the other. The parallel mechanism has the advantage of requiring only general-purpose constituent indicators. For instance, the h-neuron in the parallel case, unlike the h-neuron in the sequential case, is a general-purpose heat indicator that operates no differently than it would if it were not part of a distributed prey indicator. On the other hand, the user of a parallel indicator must be able to determine the local states of every constituent, and, as their number increases, doing so imposes an increasingly heavy burden on the user. By contrast, the sequential indicator requires specialized constituents – for instance, the h-neuron in the sequential indicator cannot be used to identify something that is warm if it is not moving – but imposes a minimal burden on the user, since the user need determine only the state of the last local indicator in the sequence. As an illustration of the difference in the burdens imposed, the user of the parallel indicator in our example must be able to distinguish the case in which the h-neuron and m-neuron both fire from three other possible cases – all the other combinations of the two neurons firing and not firing – but the user of the sequential indicator need only distinguish the case in which the h-neuron fires from the case in which it does not fire.

These differences between sequential and parallel distributed indication are due to the fact that in sequential cases each local indicator carries information not just about what properties its environment has, as in the parallel case, but also about what all the preceding local indicators indicate. The fact that the h-neuron fires in the parallel case indicates that there is something warm, but the fact that the h-neuron fires in the sequential case indicates that there is something warm *and* the m-neuron fires, and it is in virtue of indicating this that the fact that the h-neuron fires indicates that there is prey. We can think of sequential indication as a kind of distributed indication in which each indicator in the sequence of local indicators constrains the normal

conditions for the next, so that the normal conditions for each local indicator are constrained by all the preceding local indicators. In our example, if it were true that the h-neuron of the sequential distributed indicator would fire if there were a warm object, then, if the h-neuron fired, the fact that it did would fail to indicate that there was prey if the warm object were not moving. But, because of the effect of the m-neuron on the h-neuron in the context of the sequential indicator, the warm object would not be moving only if the conditions for the firing of the h-neuron were not normal.

We can also distinguish two ways in which constituents of a sequential distributed indicator can constrain the normal conditions for succeeding constituents, by having an effect on them, as in the sequential prey detector, or by having an effect on the object being tested, as in the orchid flower. In the case of sequential indicators like the flower, if an object does not have the property for which a constituent tests, the constituent (unless it is the last in the sequence) prevents it from reaching the next constituent in the sequence. But if the object does have the property for which the constituent indicator tests, the constituent (unless it is the last) allows it to reach the next constituent in the sequence, which will test the object for another property. In the flower, only those objects that pass through the opening have an opportunity to depress the pollinium lever. Thus in sequential distributed indicators like the flower, each constituent indicator (except for the last) is also a filter, and keeps objects from being tested by subsequent constituents if the object does not have the property for which the constituent tests.

By contrast, in sequential distributed indicators like the sequential prey indicator, constituent indicators have no effect on the objects they test. Instead, each constituent (except for the last) has an effect on the next constituent in the sequence. If an object has the property for which a constituent tests, the constituent (if it is not the last) gives the next constituent the opportunity to test the same object for another property. But if the object lacks the property, the next constituent stays inoperative. In the sequential prey detector, the m-neuron activates the h-

neuron if it determines that an object is moving. But unless the m-neuron determines that, it does nothing, and the h-neuron remains inhibited.

We are now in a position to contrast distributed indication with non-distributed or simple indication. The main difference between distributed indication and simple indication is that simple indication is a one-step procedure while distributed indication is a multiple-step procedure, where the steps are carried out either sequentially or in parallel depending on the kind of distributed indicator. We can think of simple indication as a matter of performing a single test on an object, or a quantity or a volume of a substance, a test that determines if what is tested is in a certain state or instantiates a certain property. It is obvious that a simple indicator can indicate multiple facts about the same object, quantity or volume. A thermometer may at one time indicate that the air around it is 70 degrees, and at another time indicate that it is 80 degrees. But for the thermometer to indicate that the air is a particular temperature, it is not required - it is not part of the normal conditions - that the volume of air be numerically the same in the two cases (that is, that the molecules that make up the volume of air in the two cases be numerically the same). By contrast, in distributed indication it is a requirement - it is among the normal conditions - that the *same* object have multiple things indicated about it. In distributed indication an indicator identifies an object as being of a certain type by determining that it instantiates a particular conjunction of properties, a conjunction of properties coextensive, when conditions are normal, with objects of that type. Thus if a simple indicator indicates two facts, what it indicates is that there is some object that is P and there is some object that is Q. But if a distributed indicator indicates two facts in a case of distributed indication, what it indicates is that there is some object that is P and Q.

3.4 Conclusion

I will conclude with some suggestions about how the theory of indication I have developed in this chapter can be extended to explain indication in people. The aim of this chapter is to develop a reductive physicalistic theory of indication that can be extended to explain distinctively human indication as a special case. I began by considering cases of indication involving simple mechanisms that indicate something in virtue of being in a single state, and then tried to show how such mechanisms can naturally be combined to form distributed indicators. In this way, I tried to show how a system can indicate a proposition of an arbitrary degree of specificity or precision by combining into a single, distributed indicator as many indicators as needed to discriminate the possible fact corresponding to the proposition from all the alternative possible facts that can obtain in the indicator's environment.

This picture suggests the following three ways in which one might expect indication in people to be different from indication in the simple mechanisms I have considered. First, it seems plausible that people have at their disposal many more potential constituent indicators than simpler indicating systems, and so can indicate much more specific propositions. Second, it also seems plausible that people have considerable flexibility in how they can combine indicators to form distributed indicators, and can combine them in novel ways. Third, it seems clear that people have feedback mechanisms that they can use to evaluate the relative efficacy of different distributed indicators formed from different combinations of constituent indicators. What all this means is that people, unlike simpler indicating systems, can *learn* how to indicate particular propositions. If a person can indicate a particular proposition, she can learn to indicate a stronger or more specific proposition, and so can improve her ability to discriminate the things that interest her from other things, and can also indicate entirely novel propositions.

Endnotes

¹ (1964) and (1976).

² Davidson (1994) 23.

³ Davidson (1963).

⁴ See Davidson (1980) 3, n. 1. Unless otherwise noted, all page numbers in subsequent citations to Davidson are for Davidson (1980).

⁵ Cf. Davidson's distinction between *causal analyses* and *causal or empirical laws* (80).

⁶ 80.

⁷ 207.

⁸ Cf. 76-7.

⁹ See Bratman (1985) for a detailed discussion of Davidson's theory of intention.

¹⁰ A reductive physicalistic strategy one might follow to explain intention starts by allowing for multiple levels of description of persons - an assumption central to functionalism. It then considers whether what plays the causal role of intention can have a level of description at which *global* states of the person's *system* of desires are explained in terms of *local* states of the *components* of the system. Models of such systems are familiar from both computer science and economics. See e.g. H. Simon (1996) and Stalnaker (1999c).

¹¹ 77.

¹² 78.

¹³ 79.

¹⁴ 79.

¹⁵ 1980.

¹⁶ 1971.

¹⁷ 1973.

¹⁸ 61. Note that where I have "if and only if," Davidson has "only if." Note also that, for Davidson, saying that an event is intentional (under a description) for an agent means the same as saying that the agent brings it about intentionally (under a description) (44; cf. xii).

¹⁹ Here I combine claims Davidson makes in (1963) and (1971).

²⁰ 4.

²¹ 4, 12-3.

²² 59. I think that in "Agency" Davidson conflates two different senses of bodily movement: (i) the *being* in motion of a part of a person's body and (ii) the *setting* in motion of a part of a person's body. As we will see, Davidson construes it in the second way when he should construe it in the first - or so I will argue.

²³ Davidson would, perhaps, prefer to say that what causes *e* is the event, the onset of *r* (12-13, 73). But it is not obvious that this would be the right thing to say in cases where what causes a person to do something is a reason she has had for some time.

²⁴ 3-4. My formulation of the definition deviates somewhat from Davidson's, and he would presumably reject my formulation, since it does not fit with his extensionalist program. But, so far as I can tell, the differences between his formulation and mine have no bearing on the problem of causal deviance.

²⁵ 3-4.

²⁶ 49.

²⁷ 52.

²⁸ 6. See also 73.

²⁹ xii.

³⁰ 46.

³¹ 61.

³² In this regard, we can distinguish Davidson's general philosophical approach to understanding thought (and thinkers) from the "heretical" view - espoused by philosophers such as Daniel Dennett and Robert Stalnaker - that we can understand thought apart from its means of expression. See Stalnaker (1999b), 1-2.

³³ 4.

³⁴ 58-9.

³⁵ (1985) lxxviii, italics added.

³⁶ 56. Davidson is quoting J. L. Austin.

³⁷ 57.

³⁸ 59.

³⁹ Davidson gives direct arguments for the specific claim that actions are single events with multiple descriptions in (1971) 57-9, and for the general claim that events are concrete entities with multiple descriptions in (1969) 181-7, but I will not consider those arguments here.

⁴⁰ Though the way of developing the single descriptum theory is mine, all the important results are Davidson's.

⁴¹ 58.

⁴² Davidson neither uses the terms "primitive description" and "non-primitive description," nor does he state a definition of primitive and non-primitive descriptions. However, the definitions are suggested by remarks he makes in (1971) (see e.g. 49 and 61). It should be noted that the distinction between primitive and non-primitive action *descriptions* is different from Arthur Danto's distinction between basic and non-basic *actions*. See Danto (1965).

⁴³ 49.

⁴⁴ For a taxonomy of different uses of "by," both causal and non-causal, see Goldman (1970) 20-48. See also Bennett (1995), Israel et al. (1993), and Lewis (1986) 256.

⁴⁵ 55-7

⁴⁶ 59.

⁴⁷ 61.

⁴⁸ Cf. 49.

⁴⁹ See the discussion of the definition schema (D) in section 1.1.2 above.

⁵⁰ Cf. 59.

⁵¹ The terminology is mine rather than Davidson's.

⁵² This definition is my attempt to make explicit what is implicit in Davidson. See 55, 59.

⁵³ Cf. 59.

⁵⁴ 78.

⁵⁵ 79.

⁵⁶ Frankfurt emphasizes the importance of actions that involve modulated movement in (1978). Davidson does mention, in a different context, extended or ongoing actions (12-3), but, in my view, fails to appreciate how widespread such actions are, and that such actions have features that make it possible to distinguish deviant from normal causal chains.

⁵⁷ Cf. Lewis (1986b), 172. My diagnosis of Davidson owes a great deal to Lewis's discussion in (1986b), 172-5.

⁵⁸ Lewis (1986c), 172-5.

⁵⁹ Cf. Lewis (1986b), 172-3.

⁶⁰ See the notion of *being intentional as an effect*, defined in (IE), in section 1.2.2.2 above.

⁶¹ As I mentioned earlier (n. 25), Davidson might prefer to say that what causes a movement in a case of action is the *onset* of a reason in the agent, rather than the *having* of the reason by the agent. Since an onset is not a process, when an onset causes something the causation must be wholesale. Thus focusing on the onset of reasons as the cause of intentional movements, and

ignoring the crucial role that, on my view, the agent's having a reason plays, might be another reason why Davidson overlooked the possibility of modulated movement.

⁶² See H. Simon (1996).

⁶³ Cf. Bratman (1987) on intention and partial plans.

⁶⁴ For the notion of one action *enabling* another, see Israel et al. (1993), n. 27.

⁶⁵ For this picture and its motivation, I am indebted to H. Simon (1996).

⁶⁶ The notion of a guiding reason is related to the notion of an intention. See Bratman (1987).

⁶⁷ In my explicit concern with the circumstances of movements in this clause of the definition and the next, I have been influenced by Israel et al. (1993).

⁶⁸ Compare Dretske's notion of a *structuring cause* in (1992).

⁶⁹ It should be noted that the existential quantifiers that occur in the previous parts of the definition – in clauses (d)(i) through (v), and in the first part of (I') – govern all of clause (d)(vi), including (d)(vi)(a) through (d).

⁷⁰ See section 1.2.1 of the last chapter.

⁷¹ See section 1.2.3 of the last chapter.

⁷² For discussions of the notion of available information, see Israel and Perry (1991b), and Stalnaker (1999c).

⁷³ Cordo and Harnad (1994).

⁷⁴ McCrea (1996), 35.

⁷⁵ McCrea (1996), 35, 38.

⁷⁶ "Stimulation of nerves in the foot produced an excitation of extensor motoneurons supplying foot muscles while concomitantly inhibiting motoneurons supplying other extensor muscles in the hind limb. Recent work...has extended these observations by showing that stimulation of the skin over the toes produced different reflexes in the foot depending upon the toe stimulated. These...reflexes seem functionally appropriate for controlling individual digits and maintaining posture....Hogbarth...showed that stimulation of the cutaneous area overlying an extensor muscle produced inhibition of other extensor muscles but excitation of the extensor muscles lying under the skin area that was stimulated." (McCrea (1996), 38).

⁷⁷ For discussions of indication, see Stalnaker (1984) and chapter 3.

⁷⁸ This is an example of *distributed* indication, which I discuss in chapter 3, section 3.3.

⁷⁹ See e.g. Dennett (1978).

⁸⁰ Stalnaker (1984).

⁸¹ Stalnaker (1984), 15. On this characterization of belief, the patterns of firing of the first two states also count as beliefs, as do changes in the shape of the foot resulting from pressure from the floor. I do not know if this is a desirable result, or if it shows that the term "dispose" that occurs in the characterization requires refinement.

⁸² *Ibid.* As in the case of the characterization of belief, this characterization of desire allows many different states of the foot to be desires. Note too that the characterizations of belief and desire allow that the same state can be both a belief and a desire – for instance, a neuron that fires only when there is pressure on a certain part of the foot might also cause the foot to move in an appropriate way. Again, I do not know if these are problems or not.

⁸³ See Perry (1993).

⁸⁴ Perhaps many implicit beliefs are self-representing.

⁸⁵ In an aberrant case, a person might w-cause one movement of her body by making another movement of her body. In that case, the claim is that she is responsible for the movement of her body that was w-caused only if it is an intentional effect of a modulated movement.

⁸⁶ Davidson (1971), 40.

⁸⁷ See definition (I') in section 1.2.3 of the last chapter

⁸⁸ See definition (IE') in section 1.2.3 of the last chapter.

⁸⁹ See e.g. Davidson (1974).

⁹⁰ For two theories of psychological states that take a view like this as their starting points, see Stalnaker (1984), and Dennett (1978).

⁹¹ See Nagel (1971). Also see Oliver Sacks (1995).

⁹² See e.g. Stalnaker (1984), 12-3, 18; Dretske (1988), chapter 3. See also Israel and Perry (1991b).

⁹³ See e.g. Stalnaker (1984), 15.

⁹⁴ My methodology and its rationale owe a great deal to Grice (1975).

⁹⁵ See e.g. Searle (1992), 245-6.

⁹⁶ See Israel and Perry (1991b), 3.

⁹⁷ Stalnaker (1984), 12-3.

⁹⁸ (1984), 12.

⁹⁹ (1991b), 4.

¹⁰⁰ Lewis (2000), 195-6.

¹⁰¹ The reasoning is as follows. Suppose that the fact that the alarm sounds indicates that the air is smoky. Then the definition implies that the alarm would sound only if the air were smoky, and also that the alarm would be in a different one of its alternative states only if the air were in a different one of its alternative states. To say the latter is to say (as I understand the counterfactual "only if") that it would be false that the alarm was in a different alternative state *if* it were false that the air was in a different alternative state. Since the states of the air form a partition, it is false that the air is in a different alternative state if and only if the air is smoky (where this biconditional is the material biconditional); and, similarly, it is false that the alarm is in a different alternative state if and only if the air is smoky. It follows that the alarm would sound if the air were smoky.

¹⁰² Not every theory of indication or information is concerned - like mine and Stalnaker's (1984), on which mine is based - with strong indication. For theories concerned only with weak indication, see Churchland and Churchland (1983), and Israel and Perry (1991b).

¹⁰³ Dretske (1981), 57; see also Israel and Perry (1991a), 154.

¹⁰⁴ See Israel and Perry (1991a), 154-7, and Israel and Perry (1991b), 15-17.

¹⁰⁵ See Lewis (2000), 196.

¹⁰⁶ For descriptions of the mechanisms of orchid fertilization, see Darwin (1877) and van der Pijl (1962).

¹⁰⁷ We can simplify the descriptions of the states of the environment of the complex indicator as follows. Consider the first state. We can formally express the assertion that an object instantiates it as:

$$(a) \quad (\exists x)((x \text{ is a gnat} \vee x \text{ is a bee}) \wedge (x \text{ is a bird} \vee x \text{ is a bee}))$$

Distributing over the conjunction in (a) we get:

$$(b) \quad (\exists x)((x \text{ is a gnat} \wedge x \text{ is a bird}) \vee (x \text{ is a gnat} \wedge x \text{ is a bee}) \vee (x \text{ is a bee} \wedge x \text{ is a bird}) \vee (x \text{ is a bee} \wedge x \text{ is a bee}))$$

In (b) there are four disjuncts, each a conjunction. Now we note that - since no object can be both a gnat and a bird, or both a gnat and a bee, or both a bee and a bird - the first three disjuncts are always false. We note, further, that we can simplify the fourth disjunct, $x \text{ is a bee} \wedge x \text{ is a bee}$, as $x \text{ is a bee}$. Accordingly, we can rewrite (b) as:

(c) $(\exists x)(x \text{ is a bee})$

We can simplify the other descriptions of the states of the environment by following similar procedures.

ⁱ⁰⁸ Churchland and Churchland (1983), 308-9.

Bibliography

- G. E. M. Anscombe. *Intention*. 1959.
- Friedrich G. Barth. *Insects and Flowers*. trans. M. A. Biederman-Thorson. 1991.
- Jon Barwise and John Perry. *Situations and Attitudes*. 1983.
- Jonathan Bennett. *Rationality*. 1989.
- Jonathan Bennett. *The Act Itself*. 1995.
- Bernard Berofsky, ed. *Free Will and Determinism*. 1966.
- Michael Bratman. *Intention, Plans and Practical Reason*. 1987.
- Michael Bratman. "Davidson's Theory of Intention." 1985. In Vermazen and Hintikka 1985.
- Michael Bratman, David Israel and Martha Pollack. 1991. "Plans and Resource-Bounded Practical Reasoning". In Cummins and Pollock 1991.
- Rodney Brooks. "A Robust Layered Control System for a Mobile Robot." 1986. *IEEE Journal of Robotics and Automation*, 2:1.
- Rodney Brooks. "Intelligence without Representation." 1991. *Artificial Intelligence*, 47.
- Roderick Chisholm. "Human Freedom and the Self." 1964. In Watson 1982b.
- Roderick Chisholm. *Person and Object*. 1976.
- Roderick Chisholm. "Agents, Causes, and Events: The Problem of Free Will." 1995. In O'Connor 1995.
- Roderick Chisholm. *A Realistic Theory of Categories*. 1996.
- Paul Churchland and Patricia Churchland. "Stalking the Wild Epistemic Engine." 1983. In Lycan 1990.
- Philip R. Cohen, Jerry Morgan, and Martha E. Polack, ed. *Intentions in Communication*. 1990.
- Randolph Clarke. "Toward a Credible Agent-Causal Account of Free Will." 1995. In O'Connor 1995.
- Paul Cordo and Stevan Harnad, eds. *Movement Control*. 1994.
- Robert Cummins and John Pollock, eds. 1991. *Philosophy and AI*. 1991.
- Arthur Danto. "Basic Actions." 1965. In White 1968.
- Charles Darwin. *The Various Contrivances by Which Orchids Are Fertilised by Insects*. 2nd ed. 1877.
- Donald Davidson. "Actions, Reasons, Causes." 1963. In Davidson 1980.
- Donald Davidson. "The Logical Form of Action Sentences." 1967a. In Davidson 1980.
- Donald Davidson. "Reply to Lemmon on Identity Conditions for Events." 1967b. In Davidson 1980.
- Donald Davidson. "The Individuation of Events." 1969. In Davidson 1980.
- Donald Davidson. "How Is Weakness of the Will Possible?" 1970a. In Davidson 1980.
- Donald Davidson. "Mental Events." 1970b. In Davidson 1980.
- Donald Davidson. "Agency." 1971. In Davidson 1980.
- Donald Davidson. "Freedom to Act." 1973. In Davidson 1980.
- Donald Davidson. "Psychology as Philosophy." 1974. In Davidson 1980.
- Donald Davidson. *Essays on Actions and Events*. 1980.
- Donald Davidson. "Davidson, Donald." 1994. In Guttenplan 1994.
- Richard Dawkins. *The Blind Watchmaker*. 1987.
- Richard Dawkins. *The Selfish Gene*. 2nd ed. 1989.
- Daniel Dennett. *Brainstorms*. 1978.
- Daniel Dennett. *Consciousness Explained*. 1991.
- Daniel Dennett. *Elbow Room*. 1993.
- Daniel Dennett. *Darwin's Dangerous Idea*. 1995.
- Fred Dretske. *Knowledge and the Flow of Information*. 1981.

- Fred Dretske. *Explaining Behavior*. 1992.
- John Fischer. *The Metaphysics of Free Will*. 1994.
- John Fischer and Mark Ravizza, eds. *Perspectives on Moral Responsibility*. 1993.
- Antony Flew. "Freedom and Human Nature." 1991. *Philosophy*. 66, 53-63.
- Jerry Fodor. *The Language of Thought*. 1975.
- Harry Frankfurt. "Freedom of the Will and the Concept of a Person." 1971. In Frankfurt 1988.
- Harry Frankfurt. "Identification and Externality." 1977. In Frankfurt 1988.
- Harry Frankfurt. "The Problem of Action." 1978. *American Philosophical Quarterly* 15. In Frankfurt 1988.
- Harry Frankfurt. *The Importance of What We Care About*. 1988.
- S. C. Gandevia and David Burke. 1994. "Does the nervous system depend on kinesthetic information to control limb movements?" In Cordo and Harnad 1994.
- Alvin Goldman. *A Theory of Human Action*. 1970.
- Paul Grice. "Method in Philosophical Psychology." 1975. *Proceedings and Addresses of the APA*, 48.
- Ned Hall. "Causation and the Price of Transitivity." 2000. *Journal of Philosophy*. 97:4, 198-222.
- Stuart Hampshire. *Freedom of the Individual*. expanded ed. 1975.
- H. L. A. Hart and Tony Honoré. *Causation in the Law*. 2nd ed. 1985.
- R. E. Hobart. "Free Will as Involving Determination and Inconceivable Without It." 1934. In Berofsky 1966.
- David Hume. *An Inquiry Concerning Human Understanding*. 1748/1955.
- David Israel and John Perry. "Information and Architecture." 1991a. In *Situation Theory and Its Applications*, vol. 2
- David Israel and John Perry. "What Is Information?" 1991b. In *Information, Language and Cognition*, ed. P. Hanson.
- David Israel, John Perry, and Syun Tutiya. "Actions, Executions, and Accomplishments." 1993. *The Philosophical Review*, 102.
- David Lewis. "Causation." 1973. In Lewis 1986c.
- David Lewis. "Events." 1986a. In Lewis 1986c.
- David Lewis. "Postscripts to 'Causation'." 1986b. In Lewis 1986c.
- David Lewis. *Philosophical Papers*, vol. II. 1986c.
- David Lewis. "Causation as Influence." 2000. *Journal of Philosophy*. 97:4, 182-197.
- William Lycan, ed. *Mind and Cognition*. 1990.
- Hugh McCann. "Volition and Basic Action." 1974. *Philosophical Review*. 83, 451-73.
- D. A. McCrea. 1994. "Can sense be made of spinal interneuron circuits?" In Cordon and Harnad 1994.
- Ernest Nagel. "Teleology Revisited." 1977. *The Journal of Philosophy*, 74:5.
- Thomas Nagel. "Brain Bisection and the Unity of Consciousness." 1971. In *Mortal Questions*.
- Thomas Nagel. "Subjective and Objective." 1979. In *Mortal Questions*.
- Thomas Nagel. *Mortal Questions*. 1979.
- Thomas Nagel. *The View From Nowhere*. 1986.
- Robert Nozick. *Philosophical Explanations*. 1981.
- Timothy O'Connor, ed. *Agents, Causes, Events*. 1995.
- John Perry. *The Essential Indexical and Other Essays*. 1993.
- John Pollock. *How to Build a Person*.
- Oliver Sacks. *The Man Who Mistook His Wife for His Hat*. 1995.
- Oliver Sacks. "A Surgeon's Life." In *An Anthropologist on Mars*. 1995.
- John Searle. "Collective Intentions and Actions." 1990. In Cohen et al. 1990.
- John Searle. *The Rediscovery of the Mind*. 1992.
- Ernest Sosa. *Causation and Conditionals*. 1975.
- Herbert Simon. *The Sciences of the Artificial*. 3rd ed. 1996.

- Robert Stalnaker. "A Theory of Conditionals." 1968. In Sosa 1975.
- Robert Stalnaker. *Inquiry*. 1984.
- Robert Stalnaker. *Context and Content*. 1999a.
- Robert Stalnaker. "Introduction." 1999b. In Stalnaker (1999a).
- Robert Stalnaker. "The Problem of Logical Omniscience, II." 1999c. In Stalnaker (1999a).
- Peter Strawson. "Freedom and Resentment." 1962. In Watson 1982b.
- L. van der Pijl and Calaway Dodson. *Orchid Flowers*. 1962.
- J. David Velleman. "Epistemic Freedom." 1989. *Pacific Philosophical Quarterly*. 70, 73-97.
- J. David Velleman. "What Happens When Someone Acts?" 1992. In Fischer and Ravizza 1993.
- Bruce Vermazen. "Negative Acts." 1985. In Vermazen and Hintikka 1985.
- Bruce Vermazen and June Merrill Hintikka, eds. *Essays on Davidson*. 1985.
- Gary Watson. "Free Agency." 1975. In Watson 1982b.
- Gary Watson. "Introduction." 1982a. In Watson 1982b.
- Gary Watson, ed. *Free Will*. 1982b.
- Gary Watson. "Free Will." 1995. In Kim and Sosa 1995.
- Alan White, ed. *The Philosophy of Action*. 1968.
- George Wilson. *The Intentionality of Human Action*. 1995.
- Ludwig Wittgenstein. *Philosophical Investigations*. 1953.
- Larry Wright. Teleological Explanations. 1976.*