Schueler, G. F., Department of Philosophy, University of New Mexico

Reasons and Purposes
Human Rationality and the Teleological Explanation of Action
Print publication date: 2003
Published to Oxford Scholarship Online: November 2003
doi:10.1093/0199250375.001.0001

Abstract: This book involves rethinking the answer to Davidson's question, "What is the relation between a reason and an action when the reason explains the action by giving the agent's reason for doing what he did?" It focuses on the thought that practical deliberation is central to explaining human action. One common version of the widely held view that explanations of actions in terms of the agent's reasons are causal explanations understands desires and beliefs as the main causal factors and says roughly that what might be called a purely causal or 'non-purposive' account of desire-belief interactions underlies the surface and (apparently) purposive or teleological explanation in terms of the agent's reasons. It is argued in this book that any such view can make no sense in the end of a common, and indeed essential, element in reasons explanations, practical reasoning itself. In the alternative account suggested here, explanations of actions in terms of the agent's reasons have an ineliminable normative element, not explicable in unadorned causal terms, which stems from the central role of practical deliberation in the genesis, and thus in the explanation, of actions. Intentional actions are always done for reasons, and the agent's reasons for doing what she did, even when there is no explicit deliberation, are whatever led her to think that this action is what she should do. So her reasons for doing what she did are intelligible only as features of her actual or possible practical deliberation, which must therefore always be at least implicitly referred to in explanations of her actions in terms of her reasons. At the same time, practical deliberation is inherently normative, both in the sense that the agent must employ evaluations in her deliberation and in the sense that her reasons are automatically open to normative criticism from herself and others. It is argued here that this requires that explanations of actions that refer essentially to the agent's deliberation have a normative element as well.

Keywords: action, causal explanation, Davidson, normative, practical deliberation, purposes, reasons, teleological explanation
Only a rational being has the power to act in accordance with his idea of laws—that is, in accordance with principles—and only so has he a will. Since reason is required in order to derive actions from laws, the will is nothing but practical reason.
Preface

In the first sentence of 'Actions, Reasons and Causes' (1963), Donald Davidson asks, 'What is the relation between a reason and an action when the reason explains the action by giving the agent's reason for doing what he did?' This is the question that this book will address. However, the answer that will emerge in the course of the discussion is very different from the one that Davidson has been widely interpreted as defending in his paper.

Davidson's answer to this question, famously, consists of two claims. The first is that what he calls 'the agent's reason' always essentially involves a desire to perform (or what he calls a 'pro attitude' toward performing) an action with a certain property and a belief that the action she performed (under the description by which she conceived it) had that property. The second claim is that this desire and belief together are the cause of the action they explain, that is, that the explanation here is a causal one.1 (Davidson's exact formulation of these two claims can be found in Section 1.2, below.)

In the years since Davidson's paper first appeared, some version of the idea that actions are explained causally in terms of the agent's 'desire-belief reasons' has come to be accepted by many, perhaps most, philosophers working in philosophy of action, philosophy of mind and connected areas, as well as by numerous social and behavioral scientists whose work bears on explanations of human action. And I think it is fair to say that even those who don't accept this answer have had a hard time agreeing on anything like a clear alternative answer to Davidson's question. So, even though deep problems have been raised about his answer to the question (for instance that it seems to commit him to epiphenomenalism2), to a large extent, Davidson's two claims still constitute for many philosophers the starting point for discussions of explanations of action (see e.g. Smith 1994, or Mele 1992). In spite of the difficulties they have encountered, they are something close to the 'received view' on these issues.

I think this is partly because of the brilliant job Davidson did in demolishing the most prominent objections to his two claims to be found in the philosophical literature at the time, perhaps most notably the objection that causes and their effects cannot stand in any sort of 'logical' relation. This and other objections to the causal view were so convincingly met that in the years since Davidson's paper the burden of proof has shifted completely. Whereas previously the philosophical climate of opinion seems rather to have favored the sort of anti-causal view advocated for instance by Ryle in The Concept of Mind (1949), currently anyone attacking the causal, desire-belief view of reasons is suspected of being guilty, at least until proven innocent, of making one or other of the mistakes exposed by Davidson. Beyond that, however, this general account of how these explanations of actions work has been understood to be consistent with, and very much in the spirit of, three widely accepted, and jointly reinforcing, doctrines.

The first, a consequence (or apparent consequence) of the 'fact-value distinction', is the view that no genuine explanation can involve in any essential way any evaluation on the part of the person doing the explaining. Genuine explanations are 'value-neutral', as it is sometimes put by the social and behavioral scientists who endorse this view. To explain an action is one thing, it is thought; to hold that it is right or wrong, or to offer any other evaluation of it, quite another.3

The second doctrine is an apparent corollary of the idea that all successful explanations of events must be causal. It is the view that so called 'belief-desire' explanations of actions must therefore be analyzed in terms of the causal interactions of these mental states (or the physical states that instantiate or 'realize' them). The thought is that genuinely causal explanations of events such as actions require reference to some other states or events as the relevant causal factors, 'the efficient causes', of which the actions are the effects. So, since

---

1 It has been argued, of course, that this second claim is itself really two distinct claims, that 'reasons are causes' and that the explanation involved is a causal one (see e.g. McGinn 1979). Hornsby (1997) argues that, while the explanations at issue are causal, it is not true—indeed, perhaps not intelligible—to hold that desires and beliefs are 'causes'. These and connected issues will be explored below.

2 See e.g. Antony (1989), Kim (1993a), or Robinson (1999).

3 I am referring here to how the belief-desire account of action explanation has been understood by many of its supporters. That Davidson himself would accept this 'value-neutral' point is not at all obvious, given his views about the role of 'radical interpretation' in explaining actions.
explanations of actions in terms of the agent's reasons are causal explanations, and a desire and belief are cited as the essential explanatory factors, the role of the desire and belief cited in the explanation must be that of the states or events that function as the efficient causes of the action in question.

There is also a third doctrine, with which Davidson's account as it is usually understood is compatible, and from which it may have gained some support. This is the doctrine that nature contains 'at bottom' no purposes, or at least none that cannot be completely explained in purely mechanical, causal terms. This is an apparent corollary of the view that all features of living organisms can be explained, directly or indirectly, by appeal to the sorts of mechanical causes that produce evolution, of which natural selection is the most important. As Richard Dawkins (1995, p. 85) colorfully puts this doctrine, 'The universe that we observe has precisely the properties we should expect if there is, at bottom, no design, no purpose, no evil and no good, nothing but pitiless indifference'.

Dawkins is discussing the sense of 'purpose' used by supporters of the argument from design, where one might for instance speak of the purpose of the cheetah's claws or the deer's antlers. But presumably if this doctrine is to be correct it has to apply to absolutely all of nature, and so to human purposes, such as Dawkins's own purpose in writing the article from which this quotation was taken.

In the pages that follow I will try to do two things. First, I will try to show that the sort of answer Davidson has been widely understood as giving to the question of how reasons explain actions is deeply problematic in ways that seem to make repair of this sort of answer impossible. While much of what I want to say on this issue builds on arguments made by others, I hope to show that, when set out fully, the case against the kind of answer Davidson has been understood as giving is decisive. (And as we will see, if I am right about this, it may well be that the answer that Davidson has been understood as giving is not in fact the answer he really gives at all.)

Second, I will try to construct an answer to Davidson's question that is sharply different from this one, one that entails, among other things, that each of the three doctrines just described is false.

Intentional actions are inherently purposive. At the same time, it is an essential, one might say 'defining', feature of intentional actions that they are done for reasons. So the agent's reason or reasons for performing some action of necessity always include the purpose or purposes for which the action was performed. So far as I can tell, these are not controversial points. It follows from them, however, that any account of how the agent's reasons explain her actions must be, or at least entail, an account of the purposiveness of the action being explained. It is the inability of the causal, desire-belief view of agents' reasons to provide such an account of the purposiveness of the action that shows most clearly where this sort of view goes wrong. At the same time, it gives an indication of how a better account can be constructed.
Acknowledgements

This book was started under a tree in my backyard in the summer of 1998. I don't know whether any good was done to the manuscript by its place of origin, but since that time it has certainly benefited from the comments, suggestions, and criticism (to say nothing of the epithets, jokes, and general bewilderment) of quite a lot of people. That fall the Philosophy Department at New Mexico State University kindly put up with my reading of an early version of part of Chapter 1. Along with their hospitality, Tim Cleveland, Richard Ketchem, Danny Socacia, and Jennifer Noonan all contributed very useful comments. A marginally more intelligible version of that same part of Chapter 1 was also delivered a bit later to participants in a conference on intentionality at the University of Oregon, organized by Bertram Malle, Louis Moses, and Dare Baldwin, where among others Bertram Malle, Joshua Knobe, and Al Mele tried valiantly to help me figure out what I was trying to say.

During the spring semester of 2000 I gave a seminar at the University of New Mexico based partly on the entire manuscript as it then was. The papers and discussion from the students in that seminar were a great help in the preparation of the next draft. Thanks are due in particular to Shelly Weinberg, Dan Gold, and Amy Lund. At about that same time Gordon Pinkham gave much of the manuscript a through editing job, which helped it greatly. Conversations about purposes and functions (and much else) with Aladdin Yaqub during that semester and at other times produced good results that showed up later at various places in the book. Other conversations with Rebecca Kukla, Amy Schmitter, Jennifer Nagel, and Sergio Tenenbaum helped me a lot in thinking about the issues discussed in Chapters 2 and 3.

Sergio Tenenbaum also read and made extensive comments on an early draft of the whole manuscript. His careful reading and insightful criticisms and suggestions were models of philosophical colleagueship. They saved me from countless mistakes and confusions. (In fact, his reading of the manuscript was so conscientious, and his comments on it so thorough, clear, and persuasive in helping with the next draft, that I think that, to be completely fair here, I have to say that in one way or another, either through omission or commission, any mistakes and confusions that remain are really his fault.)

In the fall of 2000 I was asked by Leslie Francis and Peggy Battin to read what turned out to be a version of part of Chapter 3 at the Virgil Aldrich Colloquium in the Philosophy Department of the University of Utah as the Rod P. Dixon Lecturer. The questions and comments of Michael Thompson, Michael Bratman, Elijah Mill-gram, Bruce Landesman, Nick White, and several members of the audience at that Colloquium have proved to be very helpful. Scott Sehon, Jonathan Dancy, and Dugald Owen read versions of the entire manuscript and provided extensive, and extremely useful, comments. I am very grateful to them for their efforts, which improved the book immensely. I hope they will each excuse me for not footnoting individually the numerous places where their suggestions were used.

In addition, Sandy Robbins, Trish Aragon, and especially Peter Momtchiloff deserve thanks for their help and unfailing good humor through the whole time when this book was being prepared for publication. Jessica Archibeque helped with the bibliography. Sue Hughes did the copy editing.

Finally, I owe a huge debt of gratitude to my family, Karen, Greg, and Jason Schueler. They have put up with an unconscionable amount of muddled exposition, enthusiastic unclarity, and occasional actual philosophy from me over the years. Certainly this book could never have been written without their love and support, and even if it could have been, what would have been the point?

G. F. S.

Albuquerque, New Mexico

Contents

1. Purposes, Causes, and Reasons Explanations 1
   1.1 Purposes 2
   1.2 Reasons and Causes 8
   1.3 Causes and Causal Explanations 13

2. Non-teleological Explanations of Actions 21
   2.1 The Argument for 'The Humean Theory of Motivation' 22
1 Purposes, Causes, and Reasons Explanations

Abstract: Why would anyone deny the belief-desire account of reasons explanations? It is argued here that the real question is whether the purposive, and hence apparently teleological, nature of reasons explanations in the end must be understood in non-teleological (and hence presumably purely causal or ‘mechanical’) terms.

Keywords: belief-desire account, causal explanation, causes, purposes, reasons, teleological

G. F. Schueler

It seems clear enough that intentional actions are inherently purposive; indeed, intentional human actions are paradigm examples of purposive behavior. There is always some point, aim, or goal to any intentional action. It is equally clear that our everyday explanations of actions in terms of the agent's reasons ('reasons explanations' for short) must always refer to that fact, that is to the purpose of the action, if only implicitly, on pain of not explaining the action at all. If I tell you that my reason for sprinting toward the bus stop is that the last bus leaves in five minutes, you will take this as an explanation of my action only if you assume that my purpose is to catch the last bus (or anyway that it is something involving my being there at the same time the bus is—spray painting it with graffiti perhaps). Without some such addition, my reference to the time of the last bus simply won't 'connect' in the right sort of way to what I am doing, i.e. sprinting toward the bus stop, and my action won't have been explained.

When such a purpose is identified, then at least sometimes that lets you see what I am up to; that is, it really does sometimes explain my action. It does so when the purpose mentioned really is the one for which the action was performed. The issue we will be looking at in the chapters below, at the most general level, is simply how such explanations work. So to get our bearings we should begin by looking briefly at the question of what a 'purposive' explanation is.

1 The claims that all intentional actions have some purpose and that all intentional actions are done for some reason both seem to me to be conceptually true. But nothing in what follows should depend on either claim as an assumption. So if one is doubtful of either claim one need only think of the arguments here as covering that subset of intentional actions that do have a purpose and are done for some reason. In the end, I hope, the arguments below will at least strongly suggest that both claims are true, and indeed are conceptual truths.
1.1 Purposes

An explanation is teleological, according to J. L. Mackie, ‘if it makes some essential use of the notion that something—perhaps an event, or a state, or a fact—is an end or goal to which something else is, or is seen as, a means’ (Mackie 1974, p. 279). But it is important to notice that, within this broad definition, there is a sharp distinction to be drawn between purposive explanations, strictly so called, and functional explanations.

In this respect the term ‘functionalism’, as philosophers often use it, can be misleading. In philosophy of mind, for instance, it typically refers to accounts of all mental states in terms of their ‘causal roles’, that is in terms of their causal connections to other mental and non-mental states of the person. But authors of such functionalist theories of mind usually have little or nothing to say in particular about the specific states that embody the aims or purposes of the person (or mind) being explained. Apparently they assume that these notions will be sufficiently covered in the detailed expositions of the causal relations used to explain desire, hope, fear, and other such ‘conative’ states. This is a small terminological oddity that can lead to confusion. ‘Functionalist’ theories of mental states usually give no special or separate account of an agent's purposes, simply lumping them in with all other mental states, even though in ordinary speech the term ‘purpose’ strongly overlaps with ‘function’ and purposive mental states would contrast with non-purposive ones such as belief. (The function of the coffee mug on my desk is to hold my pencils’ would, in ordinary speech, be taken to mean essentially the same as ‘The purpose of the coffee mug on my desk is to hold my pencils’.)

In biology and philosophy of biology, however, this ordinary, 'purpose' sense of 'function' is of central importance. This is at least partly because of the naturalness of speaking of the function (or purpose) of organs, traits, or behavior of animals and plants, as when someone says that the function of eyes is to see, the function of hands to grasp, or the function of green leaves to produce food for the plant to which they belong. So it is important for biologists to give an account of how eyes, hands, and green leaves can have functions or purposes, but to do so without making any reference to any person with that purpose, on pain of committing themselves to some sort of 'creationist' view, or at least to a view that allowed something to have a purpose only if some person had a purpose for it. Aristotle of course thought that things could have 'final causes' which were not the result of anyone's goals or ends. But it has been possible to explain how this could be so only since Darwin.

As they are understood in biology, an 'explanation in terms of the function' of, for instance, some trait of an organism explains the function of that trait by reference (roughly speaking) to the role the having of this trait played in the reproductive success of the ancestors of that organism. That is, such functional explanations use natural selection, usually on the basis of the adaptive benefit to the ancestors of the organism, to make sense of the assignment of functions to traits (or organs, etc.). Since natural selection is a purely mechanical process, this also means that functional explanations require no reference to any person (or to God). As Elliot Sober puts it,

The theory of evolution allows us to answer . . . two conceptual questions about function . . . . It makes sense of the idea that only some of the effects of a device are functions of the device (‘the function of the heart is to pump blood, not to make noise’). The theory also shows how assigning a function to an object requires no illicit anthropomorphism; it does not require the pretense that organisms are artifacts. (Sober 1993, p. 83)

'Anthropomorphism' in such cases would be illicit just because (and just if) the actual explanation of the trait in question is not that someone (maybe God, maybe super-beings from outer space) constructed it for that
purpose, but rather that this trait conferred an adaptive advantage of some sort on the organism's ancestors. Anthropomorphism would not be illicit, of course, if the trait in question had actually been intentionally placed in the organism by

3 Or else in terms of the causal role the trait plays in the explanation of the system in question. These two somewhat different senses of 'function' will be explained below.

4 Or perhaps that it was a 'side effect' of some trait with an adaptive advantage, a 'spandrel', as Stephen Jay Gould calls it. I don't of course mean here to take a stand on any of the various controversies about the exact role of adaptation, genetic drift, group selection, etc. in evolution.

someone with some aim in mind. That is the case, for instance, when so-called 'genetic engineering' techniques are used to produce disease resistant strains of plants by placing bits of DNA into the plants' cells. In this sort of case, although the term 'artifact' may not be the appropriate one, it seems clear enough that the explanation of that particular trait of the organism is not correctly given in terms of natural selection and adaptive benefit, but rather in terms of the aim or intention of the scientists who planned the genetic engineering that resulted in the new DNA in the organism's genes, and hence in the new trait in the plant.

So we need to distinguish 'function', in the evolutionary or adaptive sense in which natural selection can be used to make sense of the function of some trait of an organism, from 'purpose' in its ordinary sense. To see why this is important, think for a moment about how ordinary purposive explanations work. Suppose you walk into my house and see a small nail sticking out of the wall. 'What is the purpose of that?' you ask, to which the reply is 'It is for holding up the new painting we just bought.' Such an answer, I want to say, entails that someone assigned this purpose to that nail, i.e. that someone's purpose for the nail was that it hold up the painting.

I am going to take examples of this sort as paradigmatic of the sense of 'purpose' on which I want to focus, though it is perhaps worth pointing out that events, as well as objects such as nails, can have purposes in this sense. ('The purpose of that explosion was to create shock waves for our instruments to detect.') In all such cases, there are really two distinct, logically connected, uses of the term 'purpose'. We can speak of 'the purpose of something' (object, event, etc.—here the nail in that wall), and we can speak of 'someone's purpose for something'. And clearly, in this sort of case at least, assigning a purpose to some object or event entails that someone had a purpose for it. This has the consequence that any analysis or definition of 'the purpose of something' that tries to do away with this entailment must be giving a different sense of 'purpose' from this one.

To mark this point, then, I will follow the practice of many philosophers of biology and use the term 'function' for those cases where speaking of the function (or 'purpose') of a thing does not entail that someone had a purpose for it. The term 'function' will then cover at least two sorts of case: (1) causal role analyses of 'systems', such as functionalism in philosophy of mind uses for mental states,

and (2) explanations of traits or features of organisms in terms of the evolutionary history of the species in question. In the first sort of case functions are 'capacities or effects of components of systems, which are salient in the explanation of capacities of the larger system', as Peter Godfrey-Smith (1998, p. 16) puts it (following Cummins, 1975). In the second case, for something to be a function 'it must explain why the functionally characterized entity exists, and this explanation must involve some process of selection', as Godfrey-Smith says (1998, p. 16, following Wright, 1973).

The term 'purpose', on the other hand, will be used to cover cases analogous to the nail case mentioned above, where talk of the purpose of something entails that someone had that purpose for it. Roughly speaking, to talk of the purpose of something in this sense is to talk of the role the thing plays in someone's plan or project.

So one can figure out the function of a thing, in the sense of this term I am using here, either from what the thing actually is doing (its causal role) in the system of which it is a part, or from its evolutionary history, depending on which of the two sorts of function one is focusing on. Typically, in biology, where the second (evolutionary') sense of 'function' is often the relevant one, once one has discovered what it was about some trait that gave the ancestors of an organism an adaptive advantage, one has figured out the function of that trait (in the Wright sense).

But this is not so for purposes. In the sense of 'purpose' I am focusing on here, no analysis or explanation of 'the purpose of something', in terms of the microstructure of the object or its causal connections to other things or its causal history, is possible. The purpose of something, in this sense at least, is necessarily assigned to it by someone. There are presumably various ways in which this can be done, but the point is that the notion of 'the purpose of something' is relational in the sense that reference is always implicitly made to someone (or
Some group, company, etc.) who has this purpose for this thing. So trying to give an analysis or explanation in terms of causal connections or microstructure or causal history of the thing itself, for this notion of 'the purpose of this thing', would be like trying to give an analysis in terms of microstructure or causal connections or causal history of some object for the property of 'being thought of by the Mayor of San Francisco'.

Another way to see this is to notice that things can fail, even fail completely, to serve or promote their purposes. Even though the purpose of that nail in my wall may indeed be to hold up our new painting, it might in fact be far too small to do so, with the result that it never does or even could actually hold up that painting. To say that something has a certain purpose is to say in part that it should do something, or is supposed to do something (i.e. according to the role it has been assigned in some plan or project someone has), not that it actually is doing that thing or even could do it. (So 'should' here doesn't imply 'can'.) Even though we have assigned that nail the role of holding up our new painting, and that is what it should do, it might still be that it cannot fulfill this role because, unbeknownst to us, it is simply too small to bear the weight of the painting we have. Evolutionary accounts of the function of some trait, such as the color of a flower, might mimic this feature of purposes (to some extent at least), as when one says that the sexual, reproductive function of the dandelion flower is (or at least was) to attract insects for pollination, but that this function is no longer served since dandelions now reproduce asexually. But by the same token, the fact that the dandelion flower no longer plays any role in reproduction seems to justify the conclusion that it no longer has any reproductive function. That nail, however, could still have the purpose of holding up our new painting, even if there is no possibility of its doing so given the weight of the painting.

Since the purpose of a thing, in the sense I am focusing on, always comes from someone's having a purpose for it, nothing about the thing itself (or what it is doing or even could do) determines what its purpose is or even that it has a purpose. And from this it follows that there is nothing at all about what a thing is actually doing, or how it got to be where it is, or the like, that determines, or in fact is even directly logically relevant to, the question of what its purpose is, what it should be doing.

Suppose for instance that, hours later and miles away, it finally dawns on me that that little nail in the wall is far too small to hold up our new, very heavy, painting. So I consult with my family and we decide instead to use it to hold up our little plastic thermometer, which we think it can indeed do. (True, we reason, the plastic thermometer won't look nearly as nice on our living room wall as the painting; but, on the other hand, we can use the thermometer to see at a glance what the temperature of the room is, something for which the painting was no help at all.) Once we make this decision, the nail's purpose now becomes holding up our thermometer, though absolutely nothing in the nail's microstructure or causal history or causal relations to the rest of the world has changed one bit. (Presumably there are some conditions on who gets to assign purposes to what, more or less analogous to what Austin (1962) called 'felicity conditions' for speech acts such as naming. Just as I can't name your baby, or christen your ship, at least not without your permission, you can't assign a purpose to the nail in my wall. Since none of the issues to be discussed below turn on this, I won't pursue it.)

It might be worth mentioning here that the standard examples of objects with purposes—artifacts like pens or knives—are misleading in a significant way. It is true enough that artifacts have purposes only because people have purposes for them. So this is a case where 'purpose' and not merely 'function', is the appropriate term. But at the same time, it seems implausible to say that one can't simply 'see' or 'read off' the purpose of a ball-point pen, say, just by looking at it. So there would seem to be a question how this can be if, as I am arguing, there is nothing in the causal history or causal role of objects with purposes that determines what their purposes are. So it is important to be clear about how one's knowledge of the purposes of artifacts works.

One's knowledge of the purpose of, say, a ball-point pen comes not, or at least not merely, from a knowledge of its physical makeup or causal role or causal history alone, but from this plus one's knowledge of the human culture in which such pens are widely used, that is from knowledge of the fact that people have designed, constructed, and used objects of this shape and structure for a certain purpose. It is part of our knowledge of the causal history of pens that they are intentionally designed, produced, and used by people just so they will fill a certain causal role. In the terminology explained above, artifacts such as pens have the function they have (in either the causal role or causal history senses) partly because some people (primarily their designers and manufacturers but also partly their users) have constructed or adapted them to be used for certain purposes. Nothing changed (or at least nothing needed to change) in either the structure
At the same time, purposes of things are relative to plans or projects someone has for them. So it could be that my wife's purpose for that nail is to hang our new painting, and mine to hang the thermometer. Simply to ask what 'the purpose of that nail' is would be to presuppose, perhaps falsely, that it has only one purpose.

or causal history of the goose quill that was first assigned the purpose of being a pen, though presumably whoever it was who assigned it that purpose did so largely because it had the structure it did, one that fitted it for a certain role, that is for a certain function.

1.2 Reasons and Causes

Let us turn now to intentional human actions, perhaps the most common examples of 'things that have purposes', in the sense of 'purpose' I have just set out. Explanations of intentional actions in terms of the agent's reasons succeed in actually explaining the action in question, when they do explain it, partly by identifying the purpose or purposes for which the action was performed. That is the moral of the example at the beginning of this chapter of explaining my running toward the bus stop by citing my goal of catching the bus. At the same time, of course, intentional actions are events, and thus fully inside the 'causal net'. So any explanation of these actions needs to accommodate that fact as well.

There are two theses about the explanation of actions in terms of the agent's reasons that, taken together, seem to many philosophers to cover both these points, that is to keep action explanations fully inside the causal net, while at the same time displaying their essential purposiveness. These theses are thus widely held jointly. The first is the thesis that such explanations of actions must always make essential reference to some desire the agent has in combination with a belief about how to satisfy it. (I'll call this 'belief-desire thesis', BD for short.) The second is the thesis that these explanations are causal explanations (the causal thesis, or CT for short). Both theses have been held by philosophers since (and indeed including) Aristotle. In contemporary philosophy, however, these two theses have been most famously held by Donald Davidson, and since nothing I will say turns on any of the various 'fine-grained' variations of these two theses that other philosophers have adopted, it makes sense simply to stick with Davidson's formulation of them. According to him, then,

BD: R is a primary reason why an agent performed the action A under the description d only if R consists of a pro attitude of the agent towards actions with a certain property, and a belief of the agent that A, under the description d, has that property. (Davidson 1963, p. 687)

CT: A primary reason for an action is its cause. (Davidson 1963, p. 693)

In the paper from which these formulations are taken ('Actions, Reasons and Causes') Davidson argues primarily against those who wanted to deny CT because they accepted BD, that is who accepted BD but thought they saw an incompatibility between BD and CT. So the bulk of his paper consists of attempts to show that various arguments that tried to use BD as a basis for denying CT were mistaken. I want to reopen this issue, since I think that, on the only reading of BD under which it is true, it really does make CT problematic, at least if it is understood as providing an account of the explanatory force of BD, as Davidson seems to have intended.

The first problem I face in doing this, however, is in showing that there really is an issue here at all. Davidson and other philosophers who have agreed with him on this topic have done such a spectacular job of demolishing earlier mistaken arguments against CT that many of these mistakes have acquired the status of textbook examples of bad philosophical reasoning. The result is that for many philosophers it is now almost impossible to see that there is any issue here, that is to see how anyone could sensibly, and without obvious confusion, think that accepting BD made it difficult to accept CT as simply solving the problem of how reasons explanations explain. So we need to start by spending some time trying to figure out whether there is a genuine issue here, and if so what it could be. There is no doubt that many of the arguments attacked by Davidson and others in
this context really were bad ones. And I am certainly not going to try to resurrect, say, the 'logical connection' argument against CT.

The question though is whether there is any coherent position, or even any real issue, left once the confusions have been set aside. I will try to show that there really is an issue here, and I think that the best way to start is by setting out what I take to be the situation as reflected in Davidson's own discussion. Clearly, there has been a great deal of work on these issues since Davidson's paper appeared, and I am certainly not going to try to retrace the twists and turns of these debates. Instead I will set out, in a rather summary way, what I take to be their upshot. They seem to me to point toward a need for more work to be done.

Since BD seems to be accepted by both sides, CT is the place to start. Doubts about CT are doubts about whether reasons explanations are causal explanations. And of course at least some of the most prominent of Davidson's targets, such as Ryle, explicitly claimed that reasons explanations of actions are not causal (Ryle 1949). So it makes sense to begin by examining the question of what really is being claimed when one holds that reasons explanations are 'causal explanations', and in particular what has been shown and what has been left open here. It is relevant to start by focusing on an apparent, and now well known, ambiguity of CT.

In the third sentence of the paper from which the above formulations are taken, Davidson says that he will defend the position that a reasons explanation of the sort described in BD (a 'rationalization' he calls it) 'is a species of causal explanation' (Davidson 1963, p. 685). So presumably CT is intended to express that thought. But in fact CT can be read rather differently, as Davidson makes clear—and exploits—in the course of his discussion. Explanations are intensional. So one thing explains another only 'under a description'. Though the sentence 'Oedipus married Jocasta' and the sentence 'Oedipus married his mother' describe the same event, the second goes a lot further than the first in explaining Oedipus's later behavior (Mackie 1974, p. 260). But Davidson also holds that causation relates events 'no matter how described'. And CT itself merely says that a reason for an action, of the sort described in BD, 'is its cause'. So another reading of CT is possible. It can be read as merely asserting a causal relation between two things ('no matter how described'), a claim that presumably could be true even if the explanatory force of the explanation described in BD comes from something other than the causal connection referred to in CT.

So there really are two possible readings of CT here, an 'intensional' one, which interprets it as claiming that reasons explanations

of the sort described in BD are 'causal explanations', and a non-intensional one, which simply claims that the events (or whatever they are) referred to in BD are related as cause and effect. But in fact, in this context one cannot read CT as asserting that the explanation described in BD is a causal explanation, as the first reading would have it, at least not without damage to Davidson's own argument. This is because it is the non-intensional, mere-relation-between-events reading that Davidson uses to defend against a major objection to his causal thesis. The objection is that causation requires laws, and there are no genuine laws governing reasons explanations. Davidson's reply is (in essence) that this is true but irrelevant, since causation is a relation between events, however they happen to be described, and hence the only requirement is that the relevant events are governed by some laws or other, not necessarily ones using the terms of the explanation described in BD.

But this reply, though it defends CT, defends the wrong reading of it. On the resulting view, though reasons remain as causes, reasons explanations of the sort described in BD might still not be causal explanations. Worse than that, holding that the events described in BD and CT can really, in the end, only be explained under some other descriptions than 'reasons' (or 'beliefs' and 'desires') and 'action', seems to entail that 'explanations' of the sort described in BD are not really explanations at all. Though reasons will be causes, the explanatory force of reasons explanations will be completely unaccounted for. The agent's beliefs and desires will then appear to be mere 'epiphenomena' that can do no 'explanatory work', that is can contribute nothing to the explanation in which they are cited. They would be like the meanings of the words the soprano sings ('Break, O you wine glass') when it is actually the pitch and strength of her voice that
According to Davidson, 'The principle of the nomological character of causality must be read carefully: it says that when events are related as cause and effect, they have descriptions that instantiate a law. It does not say that every true singular statement of causality instantiates a law' (Davidson 1980a, p. 215).

This would be the case if for instance causal explanations really did require laws that are formulated in the same terms as the explanation, or perhaps terms that can be used to directly reduce the terms of the explanation.

Those familiar with the contemporary literature on this issue will realize that I am summing up, and taking a side on, a very vigorous debate. My point here is not to engage in that debate. Rather, as I said above, my goal is to see whether there is any coherent position 'on the other side', that is, a position that is not simply demolished by the pro-'causalist' arguments Davidson and others have given. For some of the details of this debate, see the papers in Heil and Mele (1993).

causes the wine glass to break. So if one wants to defend the idea that reasons explanations are causal explanations, something needs to be changed.

One might of course try to drop one or both of the two assumptions that led Davidson into this move in the first place, i.e. the assumption that causal explanation requires laws and the assumption that there are no genuine laws governing reasons explanations. Dropping the second of these seems hopeless for all the familiar reasons, the most telling of which is that, whatever set of beliefs and desires one tries to connect 'by causal law' to some action, it is always possible (and usually easy) to specify a further belief or desire which, if the agent has it, will lead to some other action or to none at all, thus refuting the supposed 'causal law'.

So it seems that the first assumption, that causation requires laws, will have to be the one to go if Davidson's defense is to avoid the above ambiguity, and the attendant threat of epiphenomenalism. Davidson's attempt to keep that assumption, by reading CT as referring to a mere relation between events, doesn't work, since it leaves him without an account of the explanatory force of explanations of the sort described in BD. Dropping that assumption, however, raises even more sharply the question of what the issue really is between those who hold and those who deny that reasons explanations are 'causal'. After all, if one thinks that causation requires laws, then there will be at least one very sharp issue between those who hold and those who deny that reasons explanations are causal explanations: namely, whether there really are laws in reasons explanations. If that is not the real issue, and it seems it can't be, given the obvious implausibility of holding that reasons explanations are actually lawlike (not to mention the fact that Davidson himself, famously, holds a view that entails this, the 'anomalousness of the mental'), then what is? Is there perhaps some other thesis about causation or causal explanations that divides the two sides here? To answer this question, we need to take a brief look at what it means to say that one thing causes another or that an explanation is a causal explanation.

1.3 Causes and Causal Explanations

What are 'causal' explanations? Not (or at least not merely) explanations that make essential use of the term 'cause'. Genuine causal explanations seem frequently to get along quite well without ever actually using the word 'cause' or any of its cognates.

Here is a brief (but I think perfectly accurate) causal explanation of why water comes out of the tap in my kitchen when I turn the cold water faucet handle to the right. A few miles from where I live, the city water department has pumped quite a lot of water into large storage tanks which are situated at some of the highest locations in the city, right up against the mountains. The water in these tanks is connected by underground pipes to virtually every part of the city, including my house. Since the tanks are higher than my house, there is positive water pressure in my pipes. When I turn the cold water tap to the right far enough, some holes in the...
causal priority (Mackie 1974, p. 51). It is a feature of Mackie's account of ordinary singular causal claims though that 'The distinguishing feature of causal sequence is the conjunction of necessity-in-the-circumstances and effect, or even temporally reverse causation, but as the cause being 'fixed' when the effect is not. He holds priority, since he doesn't think the ordinary concept rules out the logical possibility of simultaneity between cause and effect. Beyond this, however, he wants to add something is said to be a cause if, given the assumed background circumstances, the effect would not have happened had not the cause happened—that is, that the cause is necessary for the effect (though he also allows that we frequently think of causes as sufficient for their effects).

Of course there is plenty of 'causal terminology' in this little explanation. 'Pumped', 'turn' and so on refer to causal processes, certainly. But there is no reason to think that we would need to use the term 'cause' or any of its cognates in explaining these processes either. Just the reverse in fact, since presumably what would be wanted in the end would be (in the ideal case at least) very general physical laws and sets of initial conditions, period. Finding, in such an explanation, a sentence to the effect that one thing 'caused' another would lead us to think that there was more explaining to be done. How did it do this causing?

On the other hand, if I simply tell you that one thing has caused another, though I may have given you quite a bit of information, maybe even very useful information, I may not have done much, or even anything, to actually explain the effect in terms of the thing that caused it. It has been known for many decades that smoking causes lung cancer, but the actual mechanism by which this happens is only now being discovered. And there was quite a bit of evidence for the claim that smoking does indeed cause lung cancer (for instance, from comparison studies with matched groups of smokers and non-smokers) before there was any real idea of how the causal mechanism worked. So a straightforward way to put this situation would be to say that, while it has been known for quite awhile that smoking causes lung cancer, the actual causal explanation of how this happens is only now being worked out.

What both these points suggest is that our ordinary use of the term 'cause' would seem to be a sort of 'minimal' use, where the claim that one thing caused another amounts to nothing more than saying that the second happened because of the first, that is, that the one explains the other, period, with no information at all about how this explanation works. The word 'because', I would say, is a (or even the) minimal explanatory term in this sense, applying to anything that can be explained at all but in itself giving no content whatever to the explanatory claim. As J. L. Mackie puts this point, 'It is especially where we are inclined to ask why-questions that we will be ready to accept because-statements as answers to them, and we are particularly ready to call causes those items that can be described by clauses introduced by "because".' (Mackie 1974, p. 156). The term 'because' cites the fact that there is an explanatory story connecting two things, but by itself actually tells none of that story at all. And, as Mackie says, there is an ordinary use of 'cause' too where nothing more than that happens, though one might want to add that 'cause' seems a bit more restricted in the sorts of things to which it can apply (roughly, events, states, facts, etc.) than 'because'.

David Lewis holds that something is a causal explanation of an event if it gives any information about the causal history of that event. Given this account of 'causal explanation', then, correctly claiming that something was the cause of some event, even in the minimal sense of 'cause' explained below, would count as a causal explanation if reference to 'the cause' turned out to give any information about the causal history of the event (see Lewis 1973).

We can say '7 is the square root of 49 because 7 times 7 gives 49', but not 'The fact that 7 times 7 gives 49 causes 7 to be the square root of 49'. On the other hand, as Sergio Tenenbaum pointed out to me, mathematicians say things like 'The behavior of this number in these equations is caused by the fact that it is divisible by . . .'. Presumably this is metaphorical. Aristotle's discussion of the four kinds of cause—formal, material, efficient, and final—might be held to employ a somewhat analogous, 'minimal' sense of 'cause' to mean something like 'explanatory factor'.

Mackie's own analysis of our ordinary concept of a cause provides a well known account which is a good example of what I am calling the minimal sense of the term 'cause'. His account turns on the thought that something is said to be a cause if, given the assumed background circumstances, the effect would not have happened had not the cause happened—that is, that the cause is necessary for the effect (though he also allows that we frequently think of causes as sufficient for their effects). Beyond this, however, he wants to add the idea that in this ordinary use a cause is 'causally prior' to its effect, which he explains not as temporal priority, since he doesn't think the ordinary concept rules out the logical possibility of simultaneity between cause and effect, or even temporally reverse causation, but as the cause being 'fixed' when the effect is not. He holds that 'The distinguishing feature of causal sequence is the conjunction of necessity-in-the-circumstances with causal priority' (Mackie 1974, p. 51). It is a feature of Mackie's account of ordinary singular causal claims.
that they involve conditionals (of the form 'If C had not happened, E would not have happened'), and hence that (as Hume also held) causation is not directly observable. '[T]he distinguishing feature of causal sequences . . . is not something that can be observed in any of those sequences' (p. 54).

Whether or not one accepts Mackie's analysis of our ordinary concept of 'cause' (and of course I am giving only the barest of a bare bones description of it here), it is important to see that this analysis, like any plausible analysis of this ordinary, minimal concept, must leave completely open the question of what the ground is for the causal claim, what makes it true. That is not part of such a causal claim at all. It may be that, as many have thought, there is always some sort of regularity involved to back up any true causal claim (though Mackie himself rejects this). If so, that would ground those singular causal claims which reflected such a regularity. But in any case the question of what it is that makes the causal claim true is a completely different question from whether it is true. Saying that one thing can be explained in terms of another, which is what any such singular causal claim says, though perhaps useful and informative, is not the same as giving an explanation, at least in any but the most minimal sense.

It is more like a promissory note that says that this factor (event, state, fact) being cited will prove to be the most interesting (or important, or relevant to your purposes) in the full explanatory story, when and if it comes out. But having a promissory note, even if it is as solid as solid can be, is not the same as having whatever it is that is promised. If I promise to return the camera I borrowed from you, then, assuming I am a trustworthy person, in one sense that is as good as actually having it back. You can quite safely list it among your possessions in your will, for instance. But it is still not really the same has having it back. You can't take any actual photographs with my promise to return your camera. Similarly, saying that one thing caused another, though it might be useful in all sorts of ways (e.g. in the search for an explanation of the second thing), tells us nothing yet about how this 'causing' got done.

This minimal account of our ordinary concept of 'cause' makes sense of Davidson's statement that causes are relations between events no matter how described, which can be understood as simply making this point concisely and dramatically. To say that one thing caused another, in this minimal sense of 'cause', is really not to give a substantive explanation of the one in terms of the other: it is to claim that there is such an explanation to be had, and that these events, described in some way or other, will be referred to in that explanation. Explanations, as I said, are intensional. The content of the descriptions they contain makes all the difference. Not so in the ordinary, minimal sense of 'cause', where the only claim is that the one explains the other (on Mackie's analysis, that without the one the other would not have happened), and for this, how they are described makes no difference.

It is easy to overlook this, because the information we get in ordinary life from finding out that one thing caused another comes from

---

21 Since he also holds that the conditionals involved are non-material, which he argues are simply 'suppositions', they will frequently be neither true nor false, since the suppositions will not obtain.

22 Or 'makes it acceptable' on Mackie's account, since, given the 'suppositional' nature of such claims, they will frequently be neither true nor false, according to him (Mackie 1974, pp. 53-54).

---

the fact that the actual descriptions of the 'things' tend to be ones that interest us (compare 'The event that took place at 1:28 pm at latitude X and longitude Y caused the event that took place at 1:29 pm . . . ' with 'Your son threw a rock that broke my plate glass window'). And sometimes, in addition, it comes from the fact that picking out one specific thing as the cause serves to pin down one among a small number of possible explanatory factors. (It was the rock that broke the window, not the high note from the opera singer, not the earthquake.)

This way of understanding the claim that one thing caused another (i.e. the minimal or 'promissory note' account as exemplified by Mackie's analysis) has the advantage of demystifying claims about causation such as Davidson's. It is not that 'causation' is some mysterious, metaphysically primary, relation between events on this view. (Or at least on this account, making causal claims doesn't commit one to there being such a mysterious relation.) Saying that one thing caused another, on this account, is simply claiming that there is an explanation to be had of the second in which the first will figure importantly. At the same time, however, and for this very
reason, this raises more sharply the question of what the issue could possibly be between those who claim and those who deny that reasons explanations are 'causal explanations.'

On the minimal, 'promissory note' view of 'cause' just described, it will certainly be true that his reason for standing on his head (he thought it would impress her) caused him to do it. In this minimal sense of 'cause', at least, such a claim says nothing more than that the state (or event or thought or whatever it is) referred to by the phrase 'his reason for doing it' will figure in the explanation, whatever it turns out to be, of the event referred to by the phrase 'his standing on his head'. On Mackie's account, for instance, this will be to say that he would not have stood on his head had he not thought it would impress her (and his thinking this was 'causally prior' to his standing on his head). Since it is hard to see that anyone would deny this, it is also implausible that this is the issue that separates Davidson and other defenders of CT from those they are defending it against. If this minimal claim were all that Davidson et al. wanted to make with CT, it would make sense of why they took this to show that explanations of the sort described in BD are 'causal' (and of Davidson's own defense of CT, as I said above). But it would make deeply mysterious why anyone would want to deny it.

So we still have a puzzle here. On the minimal, promissory note, account of 'cause' (which I am claiming is simply the ordinary use of this term), reasons explanations of actions will be causal because any explanation of actions will be causal. Reasons explanations fit Mackie's analysis of 'cause' just fine. So there could hardly be an issue about whether 'reasons are causes' in this sense. And on at least one common, stronger view—the view that causal explanations, to be true, require laws (or 'lawlike connections'), as I have been arguing—it is also hard to see what the issue could be, since it seems that virtually no one who thinks 'reasons are causes' has wanted to base this view on the implausible claim that reasons explanations involve laws. Certainly Davidson, the defender of 'anomalous monism', hasn't wanted to do this.

So what is the issue? The suggestion I will defend below is that the real issue here is not whether 'reasons are causes'. In the ordinary (minimal) sense, of course they are. Rather, the issue is whether reasons explanations, which on their face always involve goals or purposes, i.e. are 'teleological' in Mackie's sense, are completely analyzable in terms of efficient causes which make no essential reference to any goals or purposes, e.g. perhaps in something like the way I argued above that functions are. It is easy to get confused here, because the term 'cause' seems often to be used by philosophers to apply to what are in fact efficient causes. So it is easy to think that in the ordinary, minimal sense 'cause' really means 'efficient cause'. If that were so, then of course showing that reasons are causes in the ordinary sense would be showing that they are efficient causes, and holding that reasons explanations are essentially teleological would entail that they are not 'causal' explanations. But, as I have been arguing, this is not so. The ordinary sense of 'cause' is a minimal, promissory note, sense which allows that reasons are causes whatever turns out to be the correct account of how reasons explanations work.

A brief look at how Mackie's analysis of our ordinary notion of cause deals with teleological explanations might help make my claim clear here. Mackie holds, as was mentioned above, that 'The distinguishing feature of causal sequence is the conjunction of necessity-in-the-circumstances with causal priority.' At the same time, Mackie's account of when an explanation is teleological is given on p. 2 above, and restated on p. 19 below.

According to him, an explanation is teleological 'if it makes some essential use of the notion that something—perhaps an event, or a state, or a fact—is an end or goal to which something else is, or is seen as, a means'. The most 'extreme' teleological claim, he says, would be where some future event or state—the 'goal'—simply determined a present or past event or state through some sort of time-reversed causal process.

Now if there were such a case, we could call the later event an end or goal, simply in virtue of its relative temporal position, rather than (as we have been doing) a time-reversed cause. And then this would be in a very clear sense a case of teleology or final causation: the 'end' would indeed be responsible for the coming about of the 'means', and any adequate explanation of these earlier-occurring items would have to refer to this end. (Mackie 1974, p. 274)

So there is no doubt that the ordinary term 'cause', as Mackie analyzes it, would apply to this sort of case, though it would simply be time-reversed efficient causation. 'The only trouble', he says, 'is that such cases never seem to occur' (Mackie 1974, pp. 274-5).

Two of the other three sorts of teleological explanation that Mackie distinguishes—animal behavior and 'goal-seeking and feedback mechanisms'—also involve efficient causation, he argues, and can be explained, without
reference to any goals, in terms of evolution or various 'feedback' mechanism; that is, they will be explained using what I have above called 'functional' explanations (Mackie 1974, pp. 277-9). So here too, though these processes are causal in the ordinary sense he explains, there is a strong reason to think that the mechanisms that explain them involve only efficient causation.

On the other hand, for the third sort of teleological explanation, involving conscious human purposive actions, Mackie argues that there is no reason to think that such underlying efficient causal explanations will be found:

[C]onscious purposive action which calls for the kind of explanatory causal account suggested does seem to involve a distinct species of efficient causation. . . . It might be that this teleological sort of account could be reduced to a more basic causal account which eliminated what I am here calling the teleological form; but there is no obvious reason why this should be possible. (Mackie 1974, p. 295)

Mackie is here using the term 'efficient causation', as he makes clear, to refer to 'the genus that covers all actual laws of working and perhaps all processes by which things come about' (p. 296). That is, he is using this to mean, in essence, what I have been calling the minimal or promissory note sense of 'cause'.

This use of the term 'efficient cause' is a bit misleading, I think, since it leaves Mackie saying that an irreducibly teleological form of causal explanation, which one would have thought would count as a paradigm of 'final causation' and hence be in sharp contrast to efficient causation, is still a 'species' of efficient causation. It would be clearer, I think, to say that, within 'the genus that covers all actual laws of working [or] . . . processes by which things come about', there might be causes of an irreducibly teleological sort, different from other, efficient, causes. This is I suppose just a linguistic quibble, but I think that ignoring it can lead to philosophers who in fact agree saying what look like very different things. For example, if I understand him, Michael Smith is making exactly the same point that I am (and that I claim Mackie is) when he says that reasons explanations may not be 'causal' at all. When he says that 'there is nothing in the debate about the theory of motivation per se that forces us to suppose that motivating reasons are causally, rather than merely teleologically, explanatory' (Smith 1994, p. 103), he is restricting 'causal' explanations to those using efficient causes; but he is certainly not at all denying that 'motivating reasons' are an essential explanatory factor in the way actions are brought about, that is that they are 'causes' in the minimal sense I have been describing.

If I am right then, the real issue here is not whether 'reasons are causes' in the ordinary (that is, minimal) sense of 'cause'. They are indeed. Rather, the issue is whether what Mackie says 'might' happen does in fact happen, that is whether reasons explanations, which certainly seem to make essential reference to goals and purposes, can—or even must—have their explanatory force explained without any such reference. This is the view of what the essential issue is here that I will try to support in the next chapter.

2 Non-Teleological Explanations of Actions

Abstract: There are well-known arguments in favour of the idea that explanations of actions are at bottom non-teleological. Michael Smith's 'direction-of-fit' argument entails this, and both Davidson and Thomas Nagel give general arguments that only causal explanations actually explain. In this chapter, it is argued that all these arguments are unsuccessful.

Keywords: action explanation, belief-desire theory, causal explanation, Davidson, direction of fit, Thomas Nagel, Michael Smith, teleological

G. F. Schueler

We have been looking for a clear understanding of whether there is really a substantive issue between, on the one side, those who follow Davidson in accepting both the belief-desire account of reasons explanations (BD) and the thesis that such explanations are causal (CT) and, on the other, those against whom Davidson and other advocates of these two claims are arguing. And I have been claiming that it is hard to find any thesis about the nature of causation, or of causal explanations, that can be plausibly held to divide the two sides here. Of course, as was noted, Ryle, in The Concept of Mind (1949), explicitly says that reasons explanations are not causal. The problem is that, on the clearest analysis of 'causal' in which this seems true, where causal explanations must involve strict laws (stated in the same terms as the explanation), Davidson and other
advocates of BD and CT explicitly accept it. So if this were someone's only reason for thinking there was a conflict between BD and CT, this opposition would indeed be based on a confusion, or at least a misunderstanding. Both sides would be in agreement on the issue of whether reasons explanations allowed strict laws stated in the terms of those explanations (i.e. would agree that they do not). At the same time, as I argued above, it is hard to see that either side would deny that CT is true under the much more minimal, 'promissory note' reading which I suggested is the ordinary sense of 'cause'.

So we still have a question concerning whether there is a genuine, coherent position opposed to the one held by the advocates of BD and CT. I think the answer to that question is 'yes'. But I think that the issue as it is often formulated, and as I have formulated it so far—the

1 And of course, as Davidson himself noted, many other philosophers have said things that seem to entail this.

issue of whether CT is true—is not really what is in dispute. Of course, merely examining a few of the numerous suggestions philosophers have made for understanding what a 'causal explanation' is, even ones that seem fairly far apart on the spectrum, is not enough to show conclusively that this is so.2 It leaves the possibility that some other notion of causal explanation is the one over which the two sides divide. But I want to explore a different possibility, namely that the real issue that divides the two sides (assuming I am right in thinking that there is one) is not simply over whether CT is true in some sense of 'cause', i.e. over whether reasons explanations are or are not 'causal' in some sense. If that is right, then it is no surprise that we have not found a way of understanding CT where the sides disagree as to its truth or falsity.

To find the real issue here, I suggest, we will need to shift gears and look at the argument offered for holding the other of the two claims Davidson makes: BD, the claim that reasons explanations of actions always and essentially involve reference to some desire, or 'pro attitude' (as Davidson calls it), of the agent (plus a belief about how to satisfy that desire). I will argue that in fact the real issue is over how to understand BD.

2.1 The Argument for 'The Humean Theory of Motivation'

Perhaps the most straightforward way of understanding BD is as a statement of what Michael Smith calls 'the Humean Theory of Motivation' (Smith 1994, chapter 4). Hume himself famously summed up his view on this issue by saying that 'Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them' (Hume 1888, p. 415). Put more prosaically, this has been taken to boil down to the thought that, since no one would intentionally or knowingly do anything that they did not want to do, there must be a desire to motivate any action that actually gets performed. Beliefs alone will never be enough. As Simon Blackburn puts it, 'Beliefs do not normally explain actions: it takes in addition a desire or concern, a caring for whatever the belief describes' (Blackburn 1998, p. 90). So the thought is this. It is a necessary condition of my doing anything intentionally that in some sense I want to do it, otherwise I wouldn't have done it intentionally. So it

2 A good bibliography for other accounts is contained in Sosa and Tooley (1993).

would seem that, if I really have no desire to do something, then I won't do it intentionally. So a desire would always seem to be required to motivate any intentional action.

This thought is frequently used as an assumption in arguments supposed to support one form of 'internalism' in ethics, the form that holds that moral or ethical commitments or beliefs must always somehow include a desire or other similar motivating factor.3 Roughly speaking the argument is that someone with a moral belief cannot just be neutral with respect to acting on that belief; someone who really has some moral belief must be inclined to act on it, at least in some circumstances. That is what it means to say that moral beliefs are 'practical' in a way that other beliefs are not. Combined with the thought that a desire is always required to motivate any intentional action, the conclusion seems to follow that 'wants' of some sort must somehow always back up or even be part of genuine moral beliefs.

Simon Blackburn uses this argument to support his version of 'expressivism' about moral thoughts. This fact (that wants are always required to actually move us), according to Blackburn, puts an 'insuperable obstacle' in the path of what he calls 'keeping ethics under the rule of Apollo', that is, roughly, accounting for moral motivation in terms of nothing but the agent's beliefs and how she reasons about them. It does this by certifying replies, in the form of arguments with a 'zigzag structure', as he puts it, to the claim that beliefs by themselves can explain actions. Each appeal to a belief to explain some act (e.g. her belief that removing the piano from his foot would
As an extra assumption the Humean Theory of Motivation, that is if we tack on (one way yen, urge) to do what was morally right (or to lessen your pain). We can get to this conclusion only if we tack on Michael Smith, the Humean Theory is entailed by the following three premises:

1. If we distinguish proper desires, completely consistent with genuine beliefs motivating actions
2. BD itself, that is as a statement of the Humean Theory of Motivation. The main contemporary argument for this theory turns on the idea that desires and beliefs have different 'directions of fit'.
3. A 'proper' desire is one where it makes sense to say that someone acted even though she had no desire at all, in this sense, to do so, as I might attend what I know will be a really boring meeting even though I had no real (i.e. 'proper') desire to go. Of course if I attended that meeting of my own free will, even if only because I regarded it as my duty to attend, then there is a sense in which I must have wanted to go. That sense of 'want' is what I am calling the 'pro attitude' sense. Here the term 'want', in this 'pro attitude' sense, simply means (something like) 'a mental state that can lead me to promote whatever it is an attitude toward'. So a proper desire, such as a craving for guacamole, might be a pro attitude, but so might a belief that something is my duty. It follows from the fact that I did something intentionally that I had a pro attitude toward doing it (or whatever I thought it would achieve); that is, it follows that there was some mental state of mine that led me to do it. That is part, at least, of what makes it an intentional action of mine. (This is why, if I did it intentionally, then I must have wanted to do it in this sense of 'want'.) But it doesn't follow that this pro attitude consisted of anything other than my belief that I had a contractual duty to do it.

Suppose that my thought that it would be morally right to lessen your pain was my reason for moving the piano off your foot. It follows that I wanted to do what was morally right (in the pro attitude sense of 'want'). But it doesn't follow at all from this that what moved me was some independent mental state, a proper desire (craving, yen, urge) to do what was morally right (or to lessen your pain). We can get to this conclusion only if we tack on as an extra assumption the Humean Theory of Motivation, that is if we tack on (one way)

This terminology comes from Schueler (1995a) pp. 29-38, following Davidson (1963). It is worth repeating that Davidson, when he uses the term 'pro attitude', seems pretty clearly to draw it broadly enough to cover any mental state that could possibly be held to move anyone, that is to be using it essentially in the sense I am. That means (as Dugald Owen pointed out to me) that Davidson himself, in advocating BD, is not in fact committed to 'the Humean Theory of Motivation', since, as I will argue below, it is only if we understand BD as referring to proper desires that it is a version of the Humean Theory.

That same distinction, I want to argue, is just as crucial for seeing the weakness in the support for this way of understanding BD itself, i.e. the assumption that such desires are always needed. That is the only way we arrive at the conclusion that moral beliefs for instance can't move us by themselves, that is aren't perfectly good pro attitudes all by themselves. But of course assuming something is not a way of arguing for it. Blackburn’s 'zigzag' argument uses BD (understood as referring to proper desires) to argue for expressivism. It is not an argument for BD itself. This is not clear, I think, unless one keeps in mind the distinction between proper desires and pro attitudes.

According to Michael Smith, the Humean Theory is entailed by the following three premises:

(a) Having a motivating reason [for some action] is, *inter alia*, having a goal;
(b) Having a goal is being in a state with which the world must fit; and
(c) Being in a state with which the world must fit is desiring. (Smith 1994, p. 116)
James Lenman spells this out as follows:

Intentional action is goal-directed: it seeks the realization of some state of the world conceived not as actual but as sought after. In this respect, to act intentionally (or merely to have some intention or desire on which one may not act) is to be in a state of mind with world-word direction of fit. A state of mind is said to have such direction of fit if the onus of match between the content of the thought involved and the world lies on the world and not the thought. Thus if I find a thought of mine with word-world direction of fit, i.e. a belief, failing to match the world, the thought is at fault and subject to revision. But if a thought with world-word direction of fit fails to match the world it is rather the world that has let us down. So if I wish to be driving toward Cambridge... but find I am not, I change not my attitude but my direction of travel and match is attained... Generally, the concern implicated in a thought with word-world direction of fit is that the thought be had only if

...its content is true. The concern implicated in a thought with world-word direction of fit is that the world be such as to match its content.

Thoughts of the latter kind—desires—contrast with those whose direction of fit is word-world: with beliefs whereby the thought content is supposed to match the world. The two are quite distinct: it is one thing to have beliefs about how things are in the world, another to care how they are. There is no belief to which we might not, in principle, be indifferent. . . .

We may take world-word direction of fit as a defining characteristic of desires thus generously understood. The central claim is then that intending to phi, in virtue of being a goal-directed state, involves having a thought of this kind: so that anyone who intentionally phi's necessarily has some desire, presumably at least the desire to phi (under some description). So, given that belief and desire are distinct, any purely cognitive understanding of the reasons that motivate us to action appears fatally incomplete. (Lehman 1996, pp. 291-292)

The argument here, I think, comes down to this. (1) Intentional actions are goal-directed; that is, they involve 'desires' in the broad sense. (2) Desires have as their defining or essential characteristic a world-to-word direction of fit. (3) Nothing with a word-to-world direction of fit (that is, no belief) can also have a world-to-word direction of fit. So, it is concluded, reasons explanations of actions must always involve desires, which is what BD says.

There is a very tempting way of reading this argument in which one might wonder why the two extra, middle premises ((2) and (3) in my version, or (b) and (c) in Smith's) are inserted here. This is because it is easy to understand the conclusion, BD, so that it obviously follows directly from the first, central, assumption that intentional actions always involve desires broadly conceived, i.e. that they are goal-directed. But I think this would miss the crucial point, that is contained in those extra premises. The real point of this argument must be to give an account of 'being goal-directed' in terms of being in one of two distinct kinds of mental or psychological state, one with a world-to-word direction of fit. That is, the thought must be that the direction of fit of the mental state more commonly called a 'desire' is the central explanatory fact of being goal-directed.

Lenman himself succumbs (or nearly succumbs) to the temptation just described when he calls world-to-word direction of fit the 'defining characteristic' of goal directedness. That comes dangerously close to merely using the term 'world-to-word direction of fit' to mean 'goal-directed'. That would merely be to give a stipulative definition about the usage of this invented term. And if that is all that is happening, the argument could not possibly yield any conclusion that couldn't be attained with the concept of being goal-directed by itself. It could not for instance, yield the conclusion that, as Lenman says, 'any purely cognitive understanding of the reasons...
that motivate us to action appears fatally incomplete'.

It is not, after all, contentious that actions involve purposes ('being goal-directed')—that is the first assumption of this argument and is not itself under dispute. And it is hardly less obvious that having a goal is different from having a garden variety factual belief. So if advocates of this direction of fit argument are doing anything beyond repeating uncontested truisms in a contorted terminology—as they must be, since their conclusion is certainly not uncontroversial—they must be proposing an explanation of why being goal-directed is different from merely having any sort of belief and is required for every action, an explanation in terms of psychological states that can have one, and only one, of two possible 'directions of fit'.

In itself, there is nothing philosophically objectionable about the 'direction of fit' terminology if it is simply used as a stylistic variant of terms such as 'belief' and 'goal-directed'. But by the same token, if that were all this argument were doing, i.e. substituting some made-up terms for some ordinary ones, then it would not really be an argument at all. It could, under this way of understanding, arrive at no conclusion that is not available without using the 'direction of fit' terminology. And in particular, it could not reach the main conclusion being sought, that beliefs alone are never enough to explain actions. So it seems to me that, if we want to understand the direction of fit argument as it is intended, that is as actually supporting the Humean Theory of Motivation (and supporting BD as a statement of this theory), as a genuine argument for it, we have to understand the phrase 'world-to-word direction of fit' as a technical term referring to some feature of mental states that provides an explanation or account of goal directedness. It cannot merely be another way of referring to goal directedness.

Another way to see that this is the way the direction of fit argument is intended comes from the fact that, while it is held to be a feature of what is called 'direction of fit' that mental states can have one or the other direction of fit, having both is explicitly ruled out. Beliefs in particular, which are paradigm examples of states with word-to-world direction of fit, are forbidden from having world-to-word direction of fit. But that is certainly not the case for pro attitudes, as this term is understood apart from this argument. As I explained it above, a pro attitude is any mental state that moves one to act. This follows Davidson's explanation. He mentions as examples of pro attitudes, along with 'desires, wantings, urges', and 'promptings', 'a great variety of moral views, aesthetic principles, economic prejudices, social conventions and public and private goals and values' (Davidson 1963, pp. 685-6). So if we were to understand 'world-to-world direction of fit' as simply applying to all the cases covered by the term 'pro attitude', it would be false that some beliefs, e.g. moral beliefs, could not have world-to-word direction of fit. And of course BD would then not look like a version of the Humean Theory of Motivation.

We can put this in terms of the distinction explained above between proper desires and pro attitudes. In order to understand the direction of fit argument as supporting the idea that beliefs by themselves can never explain actions, we cannot understand the term 'world-to-word direction of fit' as simply referring to goal directedness (purposiveness). That could at best get us to the conclusion that a pro attitude is always required for any action, and without some further argument there is nothing to say that pro attitudes (that is, mental states that can lead one to promote whatever they are attitudes toward) cannot be, say, moral beliefs. Nor can we understand 'world-to-word direction of fit' as merely referring to proper desires, since then there is a gap between the conclusion and the initial, crucial premise, that all intentional actions are goal-directed. There could still be (goal-directed) actions not motivated by proper desires. Actions performed from the thought that it is one's duty, for instance, might simply not stem from any proper desires.

So we have to understand 'world-to-word direction of fit' as referring to some feature of proper desires, which in the end, according to the argument Smith and Lenman propose, turns out to be what accounts for or explains goal directness. So in the end, according to this argument, the apparent possibility that a belief, such as a moral belief, by itself could move one to act turns out to be only apparent, not a real possibility. In short, the direction of fit argument purports to support, as a conclusion, the claim that proper desires (plus the relevant belief) are required to explain any action (that is, to support the Humean theory of Motivation) by arguing that all actions are purposive and that only mental states with a certain feature, one found only in proper desires, namely a world-to-word direction of fit, can explain purposiveness. As Smith puts it, 'Desires are . . . the only states that can constitute our motivating reasons' (Smith 1994, p. 125).

If this is right, then of course it is very important—crucial, in fact—for this argument that the explanations given of this feature of psychological states (direction of fit) be neither metaphorical nor circular, as they clearly are in
Lenman's version of this argument quoted above. If this feature of psychological states is to explain what being goal directed is, then we need some clear way of understanding direction of fit that is independent of notions such as purpose, intention, and the like. This does not happen in Lenman's discussion.

To explain world-to-word direction of fit by saying that 'the onus of match between the content of the thought involved and the world lies on the world and not the thought' is to give a misleading moral or legal metaphor, as if it is the world's job or duty to match my world-to-word psychological states. (The world has let us down' if there is no match, Lenman says.) So taking this metaphor seriously, one would get the idea that when I have a world-to-word psychological state the world somehow thereby becomes duty-bound to shape up and reorganize itself so as to match the content of my psychological state, which is obviously not what Lenman means.

Likewise, explaining the claim that '[t]he concern implicated in a thought with world-word direction of fit is that the world be such as to match its content' by a wish to get to Cambridge which leads me to change the direction in which I am driving, rather than to change my belief about it, if we take it seriously as an explanation, seems plainly circular. Wishes are just the sorts of things that are supposed to be explained by the idea that they have a world-to-word direction of fit. So, though this can serve as an example of a goal-directed state, it gets us nowhere as an actual explanation. (Anyway, if I discover that I am not going where I wish to be going, I can, after all, achieve a match between the direction in which I wish to travel and my actual direction

either by changing my direction of travel or by changing where I wish to go.)

Other philosophers, including certainly Smith himself, have tried to be more careful in explaining the idea of 'direction of fit' of psychological states. I will consider Smith's account below. In general, though, there would seem to be only two forms that such an account can take which are of relevance to the argument we are considering here. That is because the idea must be to give an account or explanation of the difference between these two sorts of mental state which isn't either circular or question-begging in the argument supporting the Humean Theory of Motivation (and so as supporting this way of understanding BD).

One form would be to give an explanation of direction of fit in terms of something like the inherent shape or structure (or other inherent features) of the two proposed psychological states themselves (where as a minimum these 'inherent' features will of course have to be fully understood without reference to concepts such as 'purpose' or 'pro attitude' that they are intended to explain). That is what the 'direction of fit' metaphor suggests, and it would be the most straightforward account that a defender of this argument for the Humean Theory could give. The other form would be to give a functionalist account in terms not of the inherent features of these states, but of their causal connections to each other. Let us look first at the sort of account that understands the two different directions of fit in terms of the inherent structure of the two psychological states.

One difficulty that arises at once is that it is hard to see why such a way of understanding 'direction of fit' wouldn't mean that there are after all laws involved in reasons explanations of actions. That is because BD would be true, on this 'inherent structure' account of mental states, only as a result of the actual interactions of these states, for instance in something like the way certain chemicals interact in chemical reactions. And then it certainly looks as if these mental state 'interactions' would themselves be describable in terms of laws, just as the chemical interactions are. It is hard to see why this wouldn't be the case, for instance, if specific mental states (types) were simply identified with brain states (types).

So on this sort of account BD itself would be a very general causal law to the effect that a certain sort of event —actions—is caused only by complexes of psychological states which include at least one state with a certain sort of inherent structure, i.e. a world-to-word direction of fit. Such an account would thus support a full-blown, 'causal law' reading of CT (rather than the minimal, mere-causal-connection or 'promissory note' reading), because it would go well beyond the minimal claim that there is some explanation of actions to be had in terms of reasons. It would propose a substantive, underlying, lawful connection between the sorts of events involved, described in terms of their inherent structures. But all this is clearly in conflict with the thought that there simply are no such psychological laws.

---


11 Such an explanation would not have to be 'reductive' in the definitional sense, of course. It could as well be a matter of explaining what such terms as 'belief' or 'desire' actually refer to, e.g. in the way early chemists figured out what molecule the term 'water' actually referred to. As J. J. C. Smart (1959) put it long ago, the identities here would be 'contingent' ones.
There are two other serious difficulties for this sort of account of these psychological states. The first is that (so far as I know) there simply are no plausible (non-circular) accounts of beliefs and desires (or 'cognitive' and 'conative' states) in terms of anything that could plausibly be called their inherent features, such as the configuration of neurons in which they are realized in the brain. Even decades ago, when type-type identity theories of mind still had fairly wide support, philosophers typically did no more than wave their hands in the direction of 'states of the central nervous system' when asked for inherent features of beliefs and other mental states. And in recent years, with defenders of type identity between mental states and brain states very few and far between, there seems not to be even this much.

The other difficulty for this sort of account is even worse, I would say, because it gives a very plausible explanation for why type-type identity theories seem so problematic. It is now widely regarded by many philosophers as a virtual truism that human mental states can only be understood as a whole, in the sense that it makes no sense to try to understand one kind of mental state, such as a desire, in isolation from most or all of the other kinds. It is this 'holistic' feature of mental states, among other things, that led to the popularity of functionalism as an account of these states. According to that view, each mental state is to be understood in terms not of its inherent structure, but of its causal interconnections with other mental and physical states of the person who has these states. Anything with the right set of interconnections is the state in question, no matter what its inherent structure turns out to be. The 'holism of the mental' entails that it is not possible to understand individual mental states, or even kinds of mental states, in terms of their inherent features. Functionalist accounts, which explain mental states in terms of the relations between different states, are consistent with this fact.

So here is another reason for thinking that the first way of explaining the direction of fit of mental states, i.e. in terms of their inherent features, is ruled out. Not only are there no actual explanations of mental states of this sort in the offing, but any non-question-begging form of such a view would seem to entail that reasons explanations of actions would indeed be describable in terms of laws, something that certainly seems false—and is in fact false if it is true that mental states can be described 'as a whole' only in some sense. That is, the holism of the mental requires reading BD much more like a conceptual truth than like the description of a causal regularity, and hence pressures us toward reading CT as about mere causal connections between events, however described (what I am calling the 'minimal' or 'promissory note' reading), rather than as saying that BD describes some sort of substantive, law-like regularity. To put it another way, understanding CT as saying that reasons explanations of the sort described in BD involve causal laws, as the 'inherent structure' analysis of 'direction of fit' seems to require, runs into the problem that it is always possible to refute any such alleged 'law' by thinking up another desire or belief the agent might have which would not lead to the action in question. And presumably this is always possible, at least partly because of the holism of the mental.

This brings us to the second general way in which supporters of reading BD as a statement of the Humean Theory of Motivation might try to explain 'direction of fit', the one that does not try to give accounts of the two 'directions' in terms of the inherent features of these mental states but somehow takes account of the holism of the mental. This would involve giving 'functionalist' accounts of beliefs and desires, or cognitive and conative mental states generally, in terms of their causal interconnections with each other and with other mental and physical states, including of course actions. As Smith puts it, on such an account 'desires are states that have a certain functional role' (Smith 1994, p. 113). Unlike an account of direction of fit in terms of the inherent structure of mental states, such a set of functionalist definitions would not, or at least not obviously, entail that reasons explanations would allow for general laws (in the terms in which they were stated), since the connections among mental states would be conceptual or definitional.

It is going to be very difficult actually to construct anything like such a full set of definitions in any detail, of course, if only because of the huge number of interconnections one would need to find and specify for each of the actual mental states being explained. The point of holism is that to do one mental state we will need to do all, or at least many. But I don't want to pursue this sort of difficulty. It is already very familiar from the many discussions of functionalism; and, who knows, perhaps supporters of the direction-of-fit argument can side-step at least some difficulties of this sort by aiming their analyses squarely at the two very general features of mental states (the two directions of fit) that they are interested in.
This is in essence what Smith tries to do by explaining the two directions of fit in terms of the different dispositions each state involves, that is (so to speak) by focusing only on the specific functional roles of states with each of the two directions of fit. According to him,

[T]he difference between beliefs and desires in terms of direction of fit can be seen to amount to a difference in the functional roles of belief and desire. Very roughly, and simplifying somewhat, it amounts, inter alia, to a difference in the counterfactual dependence of a belief that \( p \) and a desire that \( p \) on a perception with the content that not \( p \): a belief that \( p \) tends to go out of existence in the presence of a perception with the content that not \( p \), whereas a desire that \( p \) tends to endure, disposing the subject in that state to bring it about that \( p \). (Smith 1994, p. 115)

13 At the same time, though, Smith says that his account can 'remain neutral about whether desires are causes' (1994, p. 113), though he doesn't say how this can be done. Even if, as I suggested in the previous chapter, Smith can be best understood as using the term 'cause' to mean 'efficient cause', things remain a bit unclear. On the relevant sense of 'function', functional accounts of mental states will be in terms of the causal role these states play, where 'causal roles' are understood in terms of efficient causes.

14 Though it is hard to see why this wouldn't just push the issue back one stage, since it would leave the question of how there could be states of the sorts that had these conceptual interconnections unless they were realized in underlying physically describable states connected by causal laws. And since there could hardly be an infinite number of such physically describable states, why wouldn't there be laws connecting reasons and actions?

There are a number of questions one might raise about this account as is stands. For instance, it is not clear how it can deal with desires for things in the past, such as my current desire that (it be true that) I turned off the lawn sprinklers in my backyard before I left for campus this morning. There is nothing I can now do to make it the case that I did this. So it is hard to see what it could mean to say that I am somehow disposed to make it the case that I did it. More generally, there are plenty of desires that have as objects things utterly outside the agent's control. Desires about the past are only one kind of example of this larger class.15 More generally still, why should the contents of desires and beliefs be restricted to the sorts of things that can plausibly be said to be possible contents of perceptions? It is hard to see how the sort of account that Smith gives could replace 'perception' with 'belief' or some cognate without becoming plainly circular. But then, how will this account deal with imperceptible things, for instance with someone's desire that some version of Rawlsian liberalism be the morally best political theory? It might seem odd to have such a desire (though why?), but it doesn't look impossible.

But there is a bigger problem for the sort of functional account of direction of fit that Smith proposes, a problem that comes not from the implausibility of the account itself, but from the role it is supposed to play in the argument we are considering. It is important to remember that the whole direction-of-fit account is supposed to provide support for a particular way of understanding BD (for 'the Humean Theory of Motivation' as Smith calls it). And it is an essential feature of BD, on this way of understanding it, that the 'desire' referred to is a proper desire. Lenman puts this by saying that his 'concern is with the Humean claim that no purely cognitive state could, in combination with appropriate other beliefs, but with nothing else, originate a process of rational motivation' (Lenman

15 I have discussed this problem with Smith's account, as it appeared in Smith (1987), in Schueler (1991). Smith's response is at Smith (1994, pp. 208-209), where he says that a 'more accurate and fully general characterization of the functional role of desire' would be to say 'A desire that \( p \) is a state that disposes a subject to make certain sorts of bets when faced with lotteries where the outcome is inter alia that \( p \).' But, while this does at least allow Smith's account to apply to desires whose objects are outside the agent's control, it seems to me to do so only at the cost of the plausibility of this account. This 'more accurate and fully general' version doesn't seem to work when applied the garden variety cases that made the earlier version superficially plausible. Does Smith really want to say that my desire for a beer, when combined with my belief that there is one in the fridge, disposes me to make a certain sort of bet?

1996, p. 291). And, as was explained above, it is an assumption of this argument that 'Intentional action is goal directed' (p. 291). So this argument is supposed to show (or at least argue for, provide some reason to think) that the only way actions can be purposive, that is can be 'goal directed', is for the agent to be motivated by what I have called a 'proper desire', which is what Smith and Lenman propose to explain as a psychological state with world-to-word direction of fit.
But how could a functional account of direction of fit possibly do this? Any functional account of mental states, or of features of mental states, will have to take as data, as the evidential starting point on which to build the functional account, a detailed story about the roles of these states vis à vis all the other states in question. That story will be the basis on which it is decided which dispositions get built into which mental states. But this means that the truth or falsity of the Humean Theory of Motivation will have to be decided before any such functional theory gets written, that is before we can figure out which dispositions get attached to which mental states. It cannot be used to support the Humean Theory (i.e. BD understood in this way).

Consider again Blackburn's example of the person who moves the piano off someone's foot because she believes it is the morally right thing to do. If we take this at face value, it looks like this means that this belief was what moved this agent, and so that it had the disposition to produce such an action, and thus that any functional account of its 'direction of fit' will have to register this fact. But of course this would not support the 'Humean Theory' that Smith and Lenman want to support—just the reverse. It says that this belief can move someone to act without any proper desire being required. Presumably supporters of the Humean Theory will say that we should not take this example at face value. We should say that there was also a proper desire to do what is right (or to lessen the pain or the like). I will argue that this is not so, but in any case the point here is that, since this and other contentious examples are the data on which any dispositionalfunctional account of 'direction of fit' will have to be based, there is no way in which such an account can be used to support the Humean Theory. We have to decide how to describe such examples, and that includes deciding whether this Humean Theory reading of BD is correct, before we can decide which states have which dispositional connections, i.e. which directions of fit.

It is easy to miss this point, I think, because it is so easy to shift unwittingly from speaking of proper desires and the (Humean Theory of Motivation) reading of BD, which understands it as saying that a proper desire is always needed to move anyone to act, to speaking merely of pro attitudes, which refer only to whatever moved the agent to act and so could include, say, moral or other beliefs. I think this is just what happens in Smith's own defense of the Humean Theory in terms of a functional account of direction of fit. In considering what to do with mental states such as wishes, which don't seem to be exactly the same as full-blown desires, Smith says that if "desire" is not a suitably broad category of mental state to encompass all of those states with the appropriate direction of fit, then the Humean may simply define the term "pro-attitude"to mean "psycho-logical state with which the world must fit", and then claim that motivating reasons are constituted, inter alia, by pro-attitudes' (Smith 1994, p. 117). And he cites at this point Davidson's account of 'pro attitude'.

But this doesn't so much defend the Humean Theory reading of BD as utterly abandon it. As was pointed out above, Davidson explicitly includes 'moral views' under the category of 'pro attitudes'. So to adopt Smith's suggested definition here would be in essence to decide the 'data' issue represented by Blackburn's piano example against the Humean Theory reading of BD. It would entail that any accurate functional account of mental states of the sort Smith is proposing would have to allow that the belief that it was the right thing to do could, all by itself, move someone to move that piano off Blackburn's foot, which is just what the Humean Theory of Motivation denies.

So it seems to me that any attempt to defend the Humean Theory of Motivation by appealing to a functional account of the mental states that can lead to action faces an insuperable dilemma, though perhaps one that becomes clear only after we distinguish proper desires and pro attitudes. It can simply build into the account of proper desires (and nothing else) that they cause actions when combined with appropriate word-to-world states (that is, beliefs), and of course can build into the corresponding account of actions that they are caused in this way. Such a functional account will simply and flatly beg the question here. It will just assert, or assume without argument, that the contentious cases that provide apparent counterexamples to this account are decided in its favor. Alternatively, if such a functional account tries to include such states as moral beliefs in the group with world-to-word direction of fit, perhaps by being expanded so as to apply not only to proper desires but to pro attitudes generally, it simply doesn't support the Humean Theory of Motivation.

So this second, functionalist, account of direction of fit seems of no more help than the first in actually arguing for this 'Humean Theory' reading of BD. If I am right that these really are the only two general ways of explaining direction of fit, in terms of the inherent features of the mental states in question or in terms of their causal (functional) roles in the network of mental and physical states and actions to be explained, then it looks
It might be questioned whether I am right, of course. My reason for thinking that these are the only two ways possible comes from the generality of what the direction of fit argument is being called on to do, namely support the idea that proper desires are required for all intentional actions by starting from the premise that all intentional actions are goal-directed. That certainly seems to entail that the argument needs an account of the psychological state with what is called world-to-word direction of fit (that is, proper desires), which can be used to explain goal directedness in a way that does not itself make use of notions that already involve goal directedness. And that seems to restrict this argument to the two sorts of accounts I have specified, in terms of the inherent structure of mental states and in terms of their functional interconnections. These are the only two sorts of accounts of mental states that I can think of that even seem to have the potential not to involve this argument in circularity by using purposive notions as part of the explanation. If there are other sorts of accounts possible, of course, they will need to be looked at.

The point of examining the direction of fit argument, and more generally the Humean Theory it is supposed to support, was to try to discover why someone who accepted BD might think this led him to deny CT or at least question whether it accounted for the explanatory force of reasons explanations of actions. That is what Davidson and the other defenders of the idea that BD and CT can be held jointly are concerned to show. Though it may not be obvious, I think we now have the elements for answering this question. The direction of fit argument purported to support a specific ‘reading’ of BD, one that understands what Davidson calls a ‘pro attitude’ as a proper desire. This is to understand BD as stating what Smith calls ‘the Humean Theory of Motivation’. As we have seen, reading BD in this way requires direction of fit theorists to provide an account of world-to-word direction of fit (that is of ‘proper desires’) which, since it is intended to explain ‘being goal directed’, cannot make essential use of purposive or ‘goal-directed’ notions, on pain of circularity. So direction of fit theorists are forced to use some sort of causally based account, or at least an account that does not use purposive notions, which means either one that cites the inherent structure of the mental states at issue, or some sort of functionalist account.

It is this issue, I believe, that separates those who think it important to hold BD and CT jointly from those who accept BD but question CT. I have already argued that the issue is not (directly anyway) what counts as a ‘causal explanation’, or whether explanations of the sort cited in BD are causal. Rather, I believe, the issue is whether in the end explanations of the sort described in BD, reasons explanations of actions, must themselves be explained in terms of non-purposive notions of the sort that ‘direction of fit’ is supposed to be or, alternatively, whether unanalyzed purposive notions such as ‘having a goal’ or ‘having a purpose’ can do the required explanatory work. The former alternative would require that the explanatory force of reasons explanations of the sort described in BD come from something other (and presumably ‘deeper’) than the goal-directed notions employed in BD itself, hence presumably from some underlying causal regularities. That was what the direction of fit defense of the Humean Theory of Motivation was supposed to provide.

But there is another possible reading of BD, different from the one that understands it as referring to mental states such as proper desires, which could have identifying features like ‘direction of fit’, a reading different from the Humean Theory of Motivation that Smith and Lenman take as the conclusion of their argument. This is the reading of BD that both Smith and Lenman occasionally (though I think unwittingly) shift to. Instead of reading BD as referring to proper desires which, when understood, can be used to provide an explanation of the fact that actions are always goal-directed, we might read it as simply a slightly different, somewhat more detailed, way of saying the same thing: namely that actions are purposive or ‘goal-directed’. This would be to understand BD as the assumption from which the direction of fit argument as described above begins, rather than as its conclusion. This reading of BD, I will argue, though it provides no argument against CT, also provides no support for it (in anything beyond the minimal reading of it, which is uncontroversial).

It is compatible with either the truth or the falsity of any reading of CT beyond the minimal one. (I'll return to this point below.)

If this is right, then the issue is not whether some specific reading of CT (beyond the minimal one) is correct. Rather, the issue is whether there must be a (correct) version of CT which goes beyond the minimal one. That is, the issue is whether reasons explanations of actions of the sort described in BD, which certainly on their face make use of purposive concepts, must be somehow explainable or analyzable in terms that do not involve such concepts. That is what the ‘direction of fit’ terminology is supposed to do, provide a terminology with which to understand the explanatory force of reasons explanations but which itself involves no purposive concepts.
One thing that I think has made it easy to miss this issue is that it is not difficult to be misled by the fact that there really are three distinct ways to read BD. That there are at least two ways is entailed by the thought that the direction of fit argument is really an argument, rather than merely a restatement of the obvious in technical terms, since BD can be understood as either a restatement of the main premise of this argument or as its conclusion. But there is no very convenient terminology for stating even these two readings so as to make clear which is which. Neither ordinary terms such as 'want' or 'desire' nor philosophical jargon such as 'pro attitude' or 'desire very broadly understood' naturally sort out the two readings of BD that I am pointing to. So, before getting to the third possible reading of BD, it will be worthwhile restating carefully exactly what these first two readings are.

BD can be read as simply restating in a bit more detail the unexceptional and uncontroversial thought that reasons explanations give the purpose or goal for which the agent performed the action in question, which is then specified in the content of the pro attitude mentioned in BD. In this sense of the relevant term ('want', 'desire', 'pro attitude', etc.), it follows from the fact that the action was performed intentionally that the agent had some attitude of this sort. As Thomas Nagel puts it, 'That I have the appropriate desire simply follows from the fact that [some] considerations motivate me; if the likelihood that an act will promote my future happiness motivates me to perform it now, then it is appropriate to ascribe to me a desire for my own future happiness' (T. Nagel 1970, p. 29). So, to put another way, this way of reading BD simply takes it as saying that having a purpose or having a pro attitude toward something is a necessary feature of intentional action. In the now disfavored terminology, BD, on this way of reading it, is a conceptual or analytic truth. That is because, in the terminology of 'proper desires' verses 'pro attitudes' explained above, on this reading BD merely says that a pro attitude is required for any intentional action, where a 'pro attitude' is understood simply as 'a mental state that can lead one to promote whatever it is an attitude toward'.

This is in contrast to the reading of BD that I have been attributing to the defenders of the direction of fit argument, where BD, though intended to be completely general, is not analytic. This is to understand BD as a statement of what Smith calls the Humean Theory of Motivation. On this reading, the essential term of BD ('pro attitude', 'want', 'desire', etc.) refers to a proper desire and marks a feature of mental states whose explanatory role in action explanations can be set out fully only in totally non-purposive language, such as 'direction of fit' is supposed to be. BD on this reading records a general fact about actions that is itself then open to and in need of (I would say very badly in need of) further explanation. And since this feature is intended to account for the surface or apparent purposiveness of BD, it will not be, or use, a purposive notion itself. That is why, on this reading of BD, CT, which claims that reasons explanations are causal, is called on to carry real explanatory weight, to go beyond the minimal sense of 'cause'. The purposive notions in BD itself, on this reading, do not really carry that weight. They need to be explained in terms of some further feature of proper desires, which is what direction of fit is supposed to do.

That there is a third reading of BD which gets confused with the other two can be seen by looking at some of the conclusions which defenders of the teleological argument have claimed it supports. Lenman, for instance, considering the reasoning 'I know how much this person is suffering so I want to help', says: 'When we describe reasoning is, at least arguably, that one’s desires be satisfied . . . ', Lenman says (1996, p. 296). But whether or not he really wants to suggest that the direction of fit argument actually yields conclusions about what are and are not good reasons for acting (and some other things he says seem to indicate he does not), other philosophers have held not only that we must understand an agent's reasons for her action in terms of her proper desires (i.e. the second, Humean Theory, reading of BD above), but also that only desires can provide good reasons for acting, which is quite a different thought.16

So this is a third, normative, reading of BD. We can understand the term 'reason' in BD as 'good reason', and
then BD becomes a 'governing norm of practical reasoning' to the effect that, if you don't have some sort of desire or pro attitude toward doing something, then you have no good reason to do it. However plausible or implausible one finds this idea in itself, though, I hope it is clear that it is quite a different claim from either of the two embodied in the first two readings of BD already distinguished and not at all supported by any of the considerations we have been looking at so far. It is one thing to say that actions are always of necessity purposive (the first reading of BD), or that purposiveness can be explained in terms of proper desires with a feature such as world-to-word direction of fit that connects them causally to actions (the second reading), quite another to hold that one has a *good reason* to do something only if it contributes to the satisfaction of some desire or the achieving of some goal.

After all, each of the first two readings of BD, if true, presupposes that there are no actions performed for reasons in which the agent fails to have a 'pro attitude' of the sort described in BD. If there are such actions, they will constitute counterexamples to BD, understood in either of these two ways. But people sometimes act for reasons that are not good ones. So to understand BD as a norm of practical reason means that we have to allow that people sometimes actually do (or at least can do) things they do not have a pro attitude towards doing, in the sense of 'pro attitude' that is implicit in this third way of reading.

2.2 Are 'Causal' Explanations Unavoidable?

I suspect that much of the popularity of the doctrine embodied in the third, normative, reading of BD rests on confusing it with one of the other two readings. And I have been arguing that the second, Humean Theory of Motivation, reading of BD, once it is clearly distinguished from the first reading, seems at least unsupported. I don't, of course, think that anything argued above 'refutes' the second way of reading BD, only that it lets us see why advocates of this reading are committed to a less-than-minimal, and hence problematic, reading of CT. They are committed to finding the explanatory force of reasons explanations in the underlying, non-purposive, features of proper desires and their accompanying beliefs. That is at the heart of the thought that only proper desires generate purposes. Finding such a reading of CT that is plausible will not be easy. This is why I suspect that much—perhaps all—of the popularity of the second, Humean Theory, reading BD stems simply from failure to see that there is an alternative. So we need to look at the sort of explanation of action we get when we stick with the first, merely purposive, reading.

On this reading, BD is to be understood simply as a somewhat more detailed way of saying that human intentional actions are purposive, that is that the agent always has some purpose or goal in performing them, a purpose specified, or at least referred to, in the content of the description of the action under which the agent has the pro attitude toward it. Such a way of understanding BD commits its defenders to no more than the minimal reading of CT because it commits them to no further analysis or account of 'purpose'. Defenders of the direction of fit argument, since they are committed to giving an analysis or account of proper desires in terms that are not themselves purposive ('causal' terms for short, though I hope it is now clear why using this shorthand can be misleading), must hold out for some substantive, less than minimal, way of understanding CT, the claim that 'reasons are causes'. That is, on this sort of view, the purposive terms used in BD have explanatory force only because of some other, presumably 'causal', features in terms of which they can be analyzed. Proper desires, on this view, are the generators of human purposes, but they do this only by way of some underlying features which themselves are not to be explained in purposive terms. So this sort of view is committed to giving some substantive, non-minimal, and non-teleological account of what these features are and how they work.
The reading of BD I am going to be advocating here understands it as saying nothing more than that actions are performed for purposes, while leaving 'purpose' unanalyzed. Defenders of this reading are of course still committed to the minimal reading of CT. But the explanatory force of explanations of the sort described in BD will come from the fact that they are essentially purposive. There may (or may not) be a further analysis of 'purpose' or 'reason' which is plausible, but accepting the 'purposive' reading of BD by itself commits one to no position on the issue of whether there must be some further account in non-purposive terms. If I am right, therefore, in thinking that this is what really separates the two sides of the 'Are reasons causes?' debate, then, as was said, this issue has been somewhat ill-stated. It is not that one side asserts and the other side denies that reasons are causes. It is rather that one side, including defenders of the direction of fit argument, is committed to an account of the explanatory force of purposive or goal-directed explanations of intentional actions in non-purposive (hence presumably substantive causal) terms, while the other side is not.

What I have been calling the 'first' reading of BD claims that reasons explanations of actions are purposive in the sense of 'purpose' explained in Chapter 1 above, the sense in which, in order for something to have a purpose, someone must have a purpose for it. One consequence of this way of reading BD is thus that it includes an essential reference to the person who has the purpose specified in the description of the pro attitude mentioned in BD. This might seem a small point, but in fact it marks a sharp difference between the purposive and the Humean Theory readings of BD. Though BD itself

of course refers to 'the agent' who has the pro attitude and belief in question, the direction of fit account of proper desires understands the explanation involved in such a way as to make no essential use of this reference except to identify which set of beliefs and desires to describe. All the explanatory work, so to speak, on this reading of BD is supposed to get done by whatever account is given of the supposedly causal interactions of the relevant desires and beliefs themselves (or the physical states in which they are realized) in virtue of their different 'directions of fit'. (Or at least that is the hope; that is what would be so if defenders of this sort of view could find a plausible account of direction of fit.) So accepting the direction of fit account of proper desires entails an account of how reasons explanations work, which constitutes a shift to an essentially non-purposive or non-teleological form of explanation. Or rather, it constitutes a commitment to a shift to this form of explanation, a commitment to the thought that it is in the causal interactions of the relevant features of desires and beliefs that an explanation of the action in question will be found.

By contrast, according to the first, purposive, reading of BD, the explanatory mechanism of the explanations being described becomes—or rather remains—teleological, since essential use is made of 'purpose' in the sense in which, for something to have a purpose, someone must have a purpose for it. So this reading of BD understands the action being explained as something for which the agent has a purpose, which purpose itself gets specified, or at least

---

18 One might I think describe this as a shift to what Hornsby, following Dennett, calls the 'subpersonal' level (see Hornsby, 1997, esp. pp. 157-167, and Dennett, 1987, 43-68). Hornsby describes the personal level as 'a level at which mention of persons is essential, and . . . commonsense psychological explanations are indigenous to that level' (p. 161). She has an excellent discussion of some of the ramifications of the personal-subpersonal distinction, but for our purposes I think it is enough to say that the 'subpersonal level' of explanation is one at which no essential reference to a person is made and the explanations are not 'commonsense psychological explanations'. Explanations of behavior in terms of the interactions of brain states would thus be paradigm examples of subpersonal explanations. That would be one way of trying to explicate the interactions of proper desires with beliefs, but perhaps not the only such way. If it is not the only way, then perhaps the 'personal-subpersonal' terminology will not always apply to the sort of shift I am describing.

19 Of course, as long as we refer to the states in question as 'desires' and 'beliefs', we are still at the 'personal' level, since every desire or belief is always a desire or belief of someone. But at the same time, the explanatory apparatus that is being proposed, since it will operate only in terms of the interactions of these states with other such states, will itself require no such reference to a person.

---

referred to, in the description of the content of the pro attitude and belief mentioned.

What this means, I think, is that according to this reading of BD explanations of actions must involve at least two elements: first, the agent's actual reasons (as she understood them, typically), including her actual reasoning, if any, and, second, the relevant features of the agent, that is the person, who did this reasoning and had these reasons (including such things as character traits, intelligence, reasoning ability, strength of will, and so on). Unsurprisingly, these are just the things that so-called 'folk psychology' does use to explain actions (see e.g.
We may seem to have come more or less full circle here. For many philosophers at least, the whole excursion into a non-purposive analysis of BD, one part of which we have been examining, has been motivated by the thought that there is something deeply unsatisfactory, or at least very puzzling, about ordinary, so-called 'folk psychological' explanations of actions in terms of the agent's reasons, unless we regard them as explainable in non-purposive, causal terms, that is unless we give a reading of CT in such terms beyond the minimal one. That is presumably why not being committed to any reading of CT of this non-purposive sort, as what I am calling the purposive reading of BD is not, is thought to be an untenable position. If this view is correct, then whether or not the direction of fit argument is successful in explaining direction of fit so as to specify a plausible non-minimal reading of CT, that is a plausible non-teleological account of how such explanations work, we would still know in advance that some such way of understanding reasons explanations of actions must be right. So, before going on to look at how reasons explanations work if we simply stick to the purposive reading of BD, we need to face directly the question of whether they really do work at all, unless we presuppose some sort of underlying, non-purposive, causal mechanism (yet to be explained of course) between the agent's reasons and the actions they produce.

I have been arguing that each of Davidson's two claims, BD and CT, can be understood in two very different ways, in what might be called an uncontentious way and a contentious way. The uncontentious reading of BD understands it as saying nothing more than that what I have called a pro attitude is always involved in any intentional action, i.e. merely that the agent had some purpose in doing what she did. The contentious reading understands it as saying that what I have called a proper desire is required to motivate any intentional action, that beliefs alone are never enough. The uncontentious, or what I have been calling the minimal, reading of CT is simply that whatever ends up being referred to by the term 'the agent's reason' is an essential explanatory factor in explaining her action, however, this explanation ends up working. So this reading is consistent even with irreducibly teleological explanations. The contentious reading of CT understands it as saying that the agent's reason has to be the efficient cause of her action.

In Chapter 1 I argued that there is an uncontentious, 'minimal', reading of CT. On that reading, even someone who holds that reasons explanations are essentially teleological can accept it, i.e. can agree that 'reasons are causes' in a perfectly ordinary sense. In this chapter I have been arguing that the contentious reading of BD, the Humean Theory of Motivation, has not been established. At the same time, though, it could be that we can understand purposive explanations, at bottom, only as explanations in non-purposive terms. The fact that there is an uncontentious or minimal reading of CT is simply that whatever ends up being referred to by the term 'the agent's reason' is an essential explanatory factor in explaining her action, however, this explanation ends up working. So this reading is consistent even with irreducibly teleological explanations. The contentious reading of CT understands it as saying that the agent's reason has to be the efficient cause of her action. 20

The arguments for this view that I will consider here also come from Davidson (1963), who gives two positive arguments in favor of it, or perhaps two and a half. The 'half' argument is the thought that the position that explanations of actions in terms of the agent's reasons are a species of causal explanation is the ancient—and commonsense—position. That is, one might understand Davidson as holding that the causal account of reasons explanations is, so to speak, the 'default' position of common sense. This would be a burden of proof point and would make sense of the structure of the paper from which this view is taken, which as I said above consists almost entirely of refutations of attacks on the position that reasons explanations are causal explanations (as opposed to giving positive reasons for supporting this claim). The idea would be that, unless there is a successful attack on this position, it stands as acceptable without a need, so to speak, of independent argument in its support.

This is only 'half' an argument because it is never explicitly spelled out by Davidson (and so, perhaps, is not actually one he would want to make). Once it is spelled out, though, it does not look very plausible. Granting
that in some sense the commonsense position on a topic is always the 'default' position (de facto at least), it is not at all obvious that the position in question here, i.e. one that involves a substantive, non-minimal reading of CT in non-purposive terms, is the commonsense position on this issue. We certainly do give commonsense explanations of actions in terms of the agent's reasons, and sometimes at least these are explicitly her desires and beliefs. But does 'common sense' also have a position on the philosophical question of the nature of the explanation thus given?

We do of course say such things as 'She is leaving the meeting early because she wants to catch the 5.45 bus'. But it would be a mistake to think that the use of 'because' here somehow commits the speaker to an analysis of the nature of this explanation. As was argued in Section 1.3 above, 'because' simply serves to cite the important explanatory feature and is not itself 'an explanation', still less an account of how this explanation works. Its use here is merely to say that this (her desire to catch the bus) is what explains her action; that is, at most it commits 'common sense' to what I have been calling the minimal reading of CT. That would perhaps argue against any philosopher who wanted to deny that reasons were causes at all. But as an argument for, say, the Humean Theory of Motivation, so far as I can see, this 'default position' argument looks plausible only if one confuses the fact that commonsense or folk psychology does indeed explain actions in terms of agents' reasons (or tries to), and so of course entails acceptance of the minimal reading of CT, with something very different—a commitment to a further, substantive analysis in causal terms of how such explanations work.

The first (real) argument Davidson makes for the causal position is this. Discussing the view, which he attributes to followers of the later Wittgenstein, that reasons explain actions by placing them into a pattern or context in which they become intelligible, he says:

[T]alk of patterns and contexts does not answer the question of how reasons explain actions, since the relevant pattern or context contains both reason and action. One way we can explain an event is by placing it in the context of its cause; cause and effect form the sort of pattern that explains the effect, in a sense of 'explain' that we understand as well as any. If reason and action illustrate a different pattern of explanation, that pattern must be identified. (Davidson 1963, p. 692).

And two paragraphs later, after explaining Hampshire's rejection of Aristotle's claim that wanting is 'a causal factor' in producing actions, Davidson says:

But I would urge that, failing a satisfactory alternative, the best argument for a scheme like Aristotle's is that it alone promises to give an account of [what Hampshire called] the 'mysterious connection' between reasons and actions. (Davidson 1963, p. 693)

So the point here is simply that no viable alternatives have been suggested to a (substantive, non-minimal) causal account of reasons explanations. The argument would then proceed as follows. (1) Reasons explanations of actions do explain them. (2) The only viable account of how any events, including actions, are explained is in terms of cause and effect. Hence (3) reasons explanations must be a variety of cause-effect explanations. This is one of the two main positive arguments for the causal position that Davidson presents in this paper. It clearly depends on showing two things, the first of which is that the causal position is not disqualified as an account of reasons explanations in any of the ways its opponents have urged. So Davidson spends the whole final section of his paper arguing against such disqualifying claims, for example the claim that reasons couldn't be causes because there is (alleged to be) a 'logical' connection between reason and action.

This argument also depends on showing, second, that there are no other viable accounts of how reasons explain actions. That is why Davidson is concerned to refute, for instance, the 'placing-in-a-larger-pattern' account of reasons explanations. He argues that, unless this 'larger pattern' is a cause-effect pattern, it is hard to see that the action has actually been explained. But obviously, an argument of this sort, i.e. one that holds that there is really only one viable account on the table, is only as good as the actual refutations of the alleged alternative accounts that accompany it. It has no independent force of its own against any alternatives not specifically considered. It can be safely ignored, therefore, in considering the alternative account that will be presented below, which is not one that Davidson discussed.

This is not so clearly the case with the other argument in favor of a substantive reading of CT that Davidson gives, however. In discussing the justifying role of reasons, a role that causes, or at least other causes, seem not