

# In Defence of Causalism

*Philipp Huebl*

## 1. Introduction

In her essay, Maria Alvarez attacks the ‘orthodox’ causalist approach in the theory of action. For simplicity’s sake, let us identify—as Alvarez does—the orthodoxy with Donald Davidson’s action theory. Alvarez claims that the causalist approach of taking reason explanations to be elliptical causal explanations is confused. For her, reasons are facts, and thus they cannot enter into causal relations. More generally, action explanations that refer to reasons cannot be given a causal interpretation under any reading. Besides, Humean explanations of actions, which reduce the causal antecedents of actions to pairs of beliefs and desires, are overrated. Alvarez puts forward what I take to be a version of the logical connection argument by claiming that actions are the “manifestations” of mental attitudes such as desires, and hence desires cannot causally explain actions. She argues that there is a normative core notion of reason. Accordingly, she does not distinguish between different kinds of reasons, like normative reasons and motivating reasons. For her, all reason explanations are teleological explanations that make sense only in the light of the purposes or goals of the agent. Most of Alvarez’ arguments are based on evidence from English usage. She discusses occurrences of the term “reason” in our everyday parlance and examples of how we typically talk about actions.

My comment has five parts. First, I will put in perspective Davidson’s causal claim, namely the claim that reasons are causes. Second, I will argue that Alvarez fails to support her opposing view that reasons are facts. Third, I will show that her arguments against Humean action explanations go astray. Fourth, I will argue against her view that desires cannot be the causes of actions since actions are the ‘manifestations’ of desires. My final remark is on philosophical methodology based on the overall critique from the first four points. I believe that in making explicit our practice of explaining actions in everyday life and the sciences, we will rather make progress by conceptual analysis than by taking ordinary language at face value.

## 2. The Causal Claim

What exactly is the ‘causal claim’ put forth by Davidson, that is, the claim that the reason for which someone performs an action is the cause of that action? For Alvarez, the causal claim involves “the view that reasons are mental things of a kind (events, states, combinations of these) that makes them suitable candidates to be the *relata* of causal relations, and hence to be called the causes of actions.” (105)<sup>1</sup> That is not exactly Davidson’s view, but some careless formulations in his early essays invite such a reading. For Davidson, causality is a relation that holds between events only, no matter how they are described. And events are changes in the course of the world (see Davidson 1995, 272). Hence, at first sight it seems peculiar that Davidson holds that citing or indicating reasons as belief-desire pairs is citing or indicating the causes of an action, since beliefs and desires are mental *states* and thus they cannot literally be causes. However, Davidson was aware of this tension already in his essay “Action, Reasons, and Causes” (see Davidson 1963, 32 f.). It can be broken down as follows. His claim that mental states are causes must be read *with a grain of salt*. Literally only particular events, namely *changes* of those states, are causes of actions.

[I]t is *changes* in the attitudes, which are events, which are the often unmentioned causes. [...] And we can often turn a causal explanation which mentions beliefs or desires into an explanation which refers to an event or events by saying the cause of the action was the *advent* of one or both of the belief-desire pair. (Davidson 1993, 288)

However, there is a second, though unmentioned, spring for Alvarez’ quarrel with Davidson’s approach. In his later work, Davidson defines “action” by drawing a close parallel between the relation of rationalization and the relation of causation. Again, the two are of different ontological kinds: rationalization (or rational explanation) is an intensional relation between propositions, whereas causation is an extensional relation between events. Davidson never solved this problem, and there are strong arguments that it cannot be solved within his framework.<sup>2</sup>

---

1 Parenthetical references in the text are to Alvarez 2007 (in this volume).

2 For criticism, see Keil 2007, Fn 17.

### 3. Reasons as Facts?

Alvarez takes the opposite tack. She claims that reasons are what is believed or desired, not the believing or desiring of it. In other words, reasons are not mental attitudes such as states of believing or desiring, but the content of those states. The common term for the content of beliefs and desires is “proposition”. Since Alvarez also says that reasons are facts, I take her to equate true propositions with facts. At any rate, propositions and facts are abstract entities and thus reasons cannot be causes. In short, Alvarez is an *objectivist* about reasons. Her arguments are based on evidence from everyday parlance. According to Alvarez, a semantically congruent answer to the question “What was your reason for doing it?” is “she was in need” or “to help her”, but not “I believed that she was in need” or “I wanted to help her”. For Alvarez, the answers refer to facts, but not to mental states.

Some authors distinguish between two kinds of reasons: *normative* reasons and *motivating* reasons.<sup>3</sup> A normative reason says what I ought to do, given all relevant facts. A motivating reason says what is rational for me to do, given my beliefs and desires. If I know all the relevant facts and if I am rational, my normative reason becomes my motivating reason. But if I do not know all the relevant facts, I still have a motivating reason, though it may differ from my normative reason. To paraphrase an example from Derek Parfit, if I falsely believe that my hotel is on fire, I may have a motivating reason to jump out of the window, though I have no normative reason to jump. There is a controversy about what the exact relation between normative reasons and motivating reasons is. Alvarez assumes at least implicitly that the word “reason” has only one core meaning, which is displayed in everyday reason explanations. Evidence from ordinary usage is taken to answer the ontological question “what kind of thing reasons are” (104) and to describe their role in a larger framework of normativity.

According to Alvarez, our verbal practice shows that reasons are facts and that they should be separated from motivation. In fleshing out her view, she does not distinguish between examples that are commonly taken to belong to either one of the two different kinds of reasons. Let me point out three problems in this line of argument.

---

3 See Parfit 1997, 99 f. Bernard Williams uses the terms “internal reason” and “external reason”, but in a slightly different way (see Williams 1980). This distinction is also drawn by Michael Smith (see Smith 1994, 95).

First, it is only a contingent fact about the context of an utterance, whether in explaining actions we refer to our mental states or merely to the abstract propositional content of those states, or the “facts” as Alvarez would have it. Consider a courtroom case. The defendant answers to the question “What was your reason for doing it?” by uttering “to help her”. The attorney goes on asking “Did you want to help her or did somebody ask you to help her?” In this situation, it can be perfectly normal for the defendant to answer “I wanted to help her and I believed she was in danger”, or in a different situation “I did not want to help her, but my friend asked me to help her and I wanted to do him a favour.” Note, that the first sentence need not involve a Gricean implicature that in fact there was no danger at all. The courtroom example shows that in our reason explanations we *tacitly assume* or *presuppose* that propositions cited as reasons are represented by pairs of beliefs and desires. This holds even if we only mention the content of one of those states, and even if our other non-linguistic assumptions are largely left unspecified, maybe even to us.

What looks like a mere terminological issue originates from asking for the ontological category of *reasons as such*. I believe that this question already neglects the practice of explaining and rationalizing. What Davidson should have said in order to avoid confusion is something like this: only propositions can be reasons, but a proposition is not a reason unless it is actually used in reasoning. More specifically, the proposition only becomes a reason when the person has a mental attitude towards it. But this does not turn the proposition itself into a reason. Talk of reasons as such is derived from the relational use of reasons in action explanations. In other words, a fact is only a reason if it is represented as a proposition by a mental state.

Second, consider error cases. Someone went to attend a lecture, but the lecture was cancelled. Alvarez claims that the agent acted for an “apparent” reason, but she had no “real” reason. Note that here she deviates from her methodology of ordinary language analysis by drawing a technical distinction that comes close to the distinction between motivating reasons and normative reasons. In common speech, we use the term “reason” for both cases. We even use the term “reason” in cases where with some philosophical care only the term “cause” is appropriate. Apparently, the term “reason” has at least two readings as reflected in the technical terms, namely a subjective reading in “she had a reason” (motivating reason) and an objective reading in “there is a reason for her” (normative reason). In any case, what we should be interested in is not a lexicographic analysis of the term “reason”, but how the concept of reason

can be analysed. Alvarez holds that one might have an apparent reason to go to the lecture, but no reason, since reasons are facts and—as a matter of fact—the lecture was cancelled.

Apparently, this is only a redescription in different terms. Is there any difference between Alvarez' wording of "The agent had no reason, but an apparent reason" and the traditional wording of "The agent had no external reason, but a motivating reason"? At least, I can see none.

There is, however, a substantial problem about characterizing reasons without reference to mental states. I think that mental states, ushered through the front door, creep in through the back, namely counterfactually. Here is why: in order to explain what a *real reason* is, we need to say how it differs from an *apparent reason*. Why was my reason to go to the lecture only an apparent reason and not a real reason, given the lecture was cancelled? The answer is a counterfactual conditional: Had I *known* that the lecture was cancelled, I would not have gone. Knowing implies believing. So again, an explication of the concept of a real reason requires mental concepts.

Third, when we are asked for our motivating reasons for action, we are asked to answer why we did something or what motivated us to do something. Again, for Alvarez *all* reasons are abstract entities, namely facts. Facts cannot motivate us or make any change in the course of the world. As a consequence of her assumptions, reason explanations could never say or explain *why something happened*. Not even purpose explanations can, unless they are linked to causation.

But this opposes our actual practice. We cite reasons in order to explain changes in the world that we have brought about ourselves, changes that would not have occurred, had we not acted. The concept of both, motivational and normative reason is relational. A practical reason is always a reason *for* an agent *to do or want* something, and a theoretical reason is always a reason *for* an agent *to hold* a certain belief. Doing something is changing the course of the world. It is hard to see how reasons could be related to our moral and juridical practice of blame and punishment, if no causality were involved.

Alvarez separates two things in her worldview. On the one hand, there are psychological or physiological explanations of what is going on in the minds or bodies of agents when they act. On the other hand, there are reason explanations of actions in our common vernacular. But do we really talk about two different things when we explain someone's action in the light of her reasons and when we explain it in terms of her beliefs and desires?

### 3. Motivation Without Desires?

The Humean schema of action explanation takes the causal antecedents of actions to be pairs of beliefs and desires. Beliefs alone cannot motivate, they must be supplemented by some kind of positive attitude towards the action. Alvarez' arguments against Humean action explanations are problematic in several respects. Let me discuss three of them.

She considers the following example: "I am in need *can* motivate you to help me" (110), and takes this to be a case where no desire is involved. But apparently, if you are ignorant or sadistic, you would not help me, since you have no desire to do so. You might even have the opposite desire to see me suffering. Thus, the above example tacitly presupposes that you have a standing desire to help those in need. Alvarez equates *shared preferences* with *facts*. Suppose most of us human beings share the preference of helping people in need. From that it does not follow that the proposition expressed by this preference *is* a fact, namely that one should help the poor. Nor does it follow that your belief that I am in need alone can motivate you to help me.

Take another example: "The doctor prescribed him antibiotics because antibiotics are effective in fighting infections." (114) A Humean theorist might analyse this as follows: The doctor held the belief that antibiotics are effective and she wanted to help the patient. Two examples might illustrate that in action explanations we implicitly ascribe beliefs and desires to the agent. In the first case, the doctor is evil and has no inclination to help her patients. In the second case, the doctor is ignorant, for she does not believe that antibiotics are effective in fighting infections. Both examples show that the statement above makes sense only against a background of ascribing certain mental attitudes to the doctor. Alvarez believes that this view is mistaken. As a counter-analysis, she claims that in order to understand the doctor's reason we have to ascribe a goal or purpose to the doctor. I do not see how the concept of goal or purpose can shed more light on reason explanations than the concepts of belief and desire. Goals are simply desires in sheep's clothing. As desires, goals have content and they must be harboured or entertained by the agent. Wants or desires are only satisfied or fulfilled if the world changes according to their content. In John Searle's terminology, they have a mind-to-world direction of causation and a world-to-mind direction of fit (see Searle 1983, 7 f.). The same holds for goals. Introducing goals into a set of explanatory concepts is only superficially helpful for anticausalists. At first sight it seems that goals can be picked out

independently of the mental states of persons, because the term “goal” has an extensional as well as an intensional reading like “reason”. Think of a football goal. It stands there on the green even if nobody is playing the game. Yet, this metaphysical independence vanishes on closer examination. A goal is only a goal if someone has *intentionally chosen* or *constructed* it to be a goal, or if someone *wants* to reach it at some point of time. Andrew Woodfield has put this point thus:

I agree that if a goal is achieved, we say that it exists or has been actualised. But this teetering from intensional to extensional usage is a loose *façon de parler*. Intentional objects can never break free of their shackles, for they can never become real objects. What is actualised, strictly speaking, is always some action or state of affairs that *matches* the goal by satisfying a goal-description. (Woodfield 1976, 211)

In criticising Alfred Mele, Alvarez equates two action explanations schemas, the *subsumption-reading* and the *means-end-reading*. She claims

So, if my reason for singing were that I wanted to sing, then it should make sense to say that I sang in order to sing (compare: ‘My reason for driving was that I wanted to get there in time’—‘I drove in order to get there in time’). (111)

The example about driving has a means-end-reading. The agent wants to be somewhere in time, and believes that driving is the mean to get there in time. However, the example about singing is different. The agent was singing because she had a desire to sing. The particular action can be subsumed under the action type ‘singing’. In this case there is no means-end-reading. Some actions are *basic* in the sense that we do not have to perform another action in order to perform the basic action. For Davidson, whether an action is basic depends entirely on the description under which it is picked out. I think this assumption needs some qualification. In my view, whether an action is basic depends on the content of the intention it is performed with. I may intend to move my fingers one by one according to the musical score. But if I know how to play the sonata, I can intend to play it as a whole. In the first case, every single finger movement may be a basic action, and in the second case, playing the whole sonata may be a basic action. My reason for playing a sonata may be to entertain others, to make money, or to fulfil a commitment. But I may also play it just for the sake of it, and then my reason is my plain desire to play it. The means-end reading is not a necessary condition for attributing reasons to an agent. And consequently, a causalist is not forced to claim that all reason explanations fit the

instrumental means-end-schema. However, since sometimes the term “Humean explanations” is used exclusively for the instrumental kind, it is worth mentioning that at least Davidson is not a Humean in the narrow sense.

#### 4. Actions as Manifestations of Desires?

Alvarez admits that she finds an argument compelling that goes something like this: the conceptual connection between desires and actions is so close that there cannot be a causal connection. In her words “Such actions are the *manifestations*, rather than the effects, of the agent’s wanting something.” (117 f.). I take this to be a version of the logical connection argument, but since Alvarez explicitly denies this, I am happy to call it the ‘conceptual connection argument’. I am not convinced either way. From the fact, that we use behavioural evidence in attributing mental states to other persons, it does not follow that we cannot have desires before putting them to action. One can have the desire to climb Mount Everest without ever getting the chance to satisfy it, for instance if one serves a life sentence in prison. Besides, the same independence occurs in the reverse direction: Some actions satisfy or fulfil some of our desires, but are not related to them in any direct sense.

It is difficult to tell what Alvarez means by her remark that “[...] in advance of the relevant action, it is partly an open question whether they [the agents] indeed have the goal, or the extent to which they have it—it is an open question not because we cannot *yet know* it but because the answer is still indeterminate.” (118) An open question to whom? To us, the observers of the actions of others? Or to us, the performers of our own actions? I believe Alvarez merges two kinds of uncertainties. On the one hand, there can be *uncertainty in attributing beliefs and desires* to other persons. We do this solely on behavioural evidence, since there is no other kind of evidence. Needless to say, this enterprise is not safe of error. On the other hand, we can be *uncertain about the status of one of our own desires* among the set of others. Alvarez seems to have the second kind of uncertainty in mind. But then it is not a point about desires in general and whether they are ‘constituted’ (whatever exactly that means) by actions, but rather a point about our biological species. Although we as human beings have largely first-person authority about our own mental states, we do not always have perfect access to them. And even when we have, sometimes we misrepresent the world. However, we could easily



imagine higher developed rational beings that always have this kind of access and that almost never misrepresent the world. Yet, even given our limited capacities, uncertainty does not render the relation of desires and actions conceptual.

Let me flesh out the last remark: Sometimes we do not know what we prefer, until we have tried the options. Consider examples involving preferences for chocolate bars. I have two bars in front of me, a bitter sweet chocolate bar and a milk chocolate bar. What are my preferences? There are four possibilities: I prefer  $x$  to  $y$ , I prefer  $y$  to  $x$ , I like  $x$  and  $y$  to the same degree, or I lack preferences concerning them, for instance if I do not even know what chocolate is. A misrepresentation of my preferences can have at least two sources. On the one hand, in the first two cases of asymmetric preferences, my desire is specific enough to decide between the two bars. Yet, I might be mistaken about how one or each of them really tastes like. In other words, I might be mistaken about how I represent the world. For example, I believe that this bar has the taste I prefer most, but in fact it has a different taste. And by comparison, I find out that I like the other chocolate bar better. On the other hand, in the case of symmetric preferences my desire is not specific enough to decide between the two bars. I just do not know myself well enough in this very moment. Either, I do not have perfect access to my first order desires. Technically speaking, I may need to develop a higher-order mental state to represent my first-order preference. Or, I may need to acquire a preference hierarchy by *actually tasting* the bars. The same holds for cases, where I have no preferences at all.

In the cases discussed, I have to try the chocolate bars in order to 'manifest' my desire, at least in one sense of the term "manifest". And by tasting the bars I might find out something new about myself. In a loose sense of the term "open question", before tasting the chocolate bars it was an open question for me what preferences I had. In any event, most cases involving desires and actions do not involve misrepresentation or underdetermination of preferences. I know what I want and I could tell you without first observing my own actions. And even in cases of misrepresentation: my desire is what *makes me act* in the first place. It is part of my reason for *actually tasting* the chocolate.

There is another way of reading the passage, namely as a point about metaphysics and not epistemology. Desires are mental states that represent the world via their content. The same desire can be fulfilled in an indefinite number of ways, because there are blank positions in its content that can be occupied in many ways by the world without

affecting the fulfilment of the desire as a whole. The world itself is ontological denser, so to speak, than its representation.<sup>4</sup> However, this metaphysical fact does not make having the desire or goal indeterminate, as Alvarez believes. It concerns something entirely different, namely the actual ways of fulfilling a desire, or of reaching a goal. These different permutations leave the content of the desire untouched.

Here is another argument. Consider desires and intentions. Both are kinds of motivational or pro-attitudes in having the world-to-mind direction of fit: they are satisfied or fulfilled, if the world changes according to their content. Intentions, however, concern only changes that can be brought about by actions, whereas desires include changes beyond the agent's control. At least if I am rational, I cannot have the intention that the sun may shine tomorrow or to live forever. But I can have a desire, or want, or wish with the same content, even if I do not believe that I could *do anything* to make it happen. Apparently, those desires 'manifest' themselves somehow in speech, but there are no *specific* actions that are 'partly constitutive' for entertaining those desires.

### 5. Conceptual Explication as Philosophical Methodology

I take philosophy to be the science that deals with the most basic concepts and how they are related to one another.<sup>5</sup> A philosophical action theory reconstructs explicit and implicit assumptions about actions explanation in everyday life and the sciences by relating the basic concepts to one another, which underlie this practice. Alvarez adheres to a narrower approach. She refers to overt linguistic evidence from ordinary usage of English. According to her, we typically say "the reason was that it was raining", but not "the reason was that I believed that it was raining". Alvarez draws substantial conclusions from that finding. For her, since we do not use the words "belief" and "desire" in any direct paraphrase, we have no grounds to assume that the concepts of belief and desire play a role in reason explanations. Yet this misrepresents the whole point about conceptual analysis. The causalist's claim is that reason explanations are *implicit*, or *elliptical*, or *metonymic* causal explanations. In order to make our largely tacit practice of explaining actions explicit, we need to introduce two basic mental attitudes, belief and desire, and take changes

4 See Keil 2007, 81–83.

5 What Peter Strawson calls "connective analysis"; see Strawson 1992, chap. 2.

of pairs of them as causes. In other words, talk about actions rests on some of our most basic concepts such as “belief”, “desire”, “cause”, and “event”.

I do not see the point in taking ordinary language examples at face value, since when we talk, only few of our assumptions and only few ascriptions of mental attitudes have a reflex in our words. Hence, as long as Alvarez does not provide a theory of language comprehension, which specifies how overt linguistic data and implicit linguistic and other assumptions interplay, it is futile to regard examples from a single natural language as sacrosanct. Especially, as long as these generalisations are not shown to hold for all (possible) languages.

In two cases, Alvarez even departs from her own course. First, when she allows technical terms to come into play, like “real reason” and “apparent reason”. Evidently, non-philosophical speakers often use the same word “reason” for both cases, but can be made aware of the difference. So shall we take their initial use of words seriously, or their use after they have learned the distinction? Second, Alvarez says “[t]o explain an action by reference to the agent’s reasons is to introduce (perhaps implicitly) the notion of a goal, in particular the goal or end for the sake of which the action was done.” (120) When Alvarez allows ‘implicit introduction’ of concepts, she assumes precisely what causalists already said about beliefs and desires. Does her analysis in terms of the concept of goal have any advantage over the orthodoxy? I do not think it has. To say someone has a goal, presupposes that she *wants* to reach this goal. In other words, the concept of goal presupposes the concepts of intention or desire. In the end, one has to assume that the goal is somehow mentally represented by the agent. Something similar applies to downplaying the role of Humean explanations. The defenders never claimed that they often occur in their genuine form. Humeans never made a point about relative frequency of schemas of explanations, but a point about the presupposed form and the most basic concepts of actions explanations.

### Bibliography

- Alvarez, Maria (2007): “The Causalist/Anti-causalist Debate in the Theory of Action: What it is and Why it Matters” (this volume).
- Davidson, Donald (1963): “Action, Reason, and Causes”, reprinted in: Davidson, Donald (1980): *Essays on Actions and Events*, Oxford: Oxford University Press.

- (1993): “Replies”, in: Stoecker, Ralf (ed.): *Reflecting Davidson. Donald Davidson Responding to an International Forum of Philosophers*, Berlin/New York: De Gruyter.
- (1995): “Laws and Cause”, in: *Dialectica* 49, 263–279.
- Keil, Geert (2007): “What Do Deviant Causal Chains Deviate From?” in Lumer, Christoph/Nannini, Sandro (eds.): *Intentionality, Deliberation and Autonomy: The Action-Theoretic Basis of Practical Philosophy*, Aldershot: Ashgate (forthcoming).
- Parfit, Derek (1997): “Reasons and Motivation”, in: *The Aristotelian Society, Supplementary Volume* 77, 99–130.
- Searle, John (1983): *Intentionality*, Cambridge: Cambridge University Press.
- Smith, Michael (1994): *The Moral Problem*, Oxford: Blackwell Publishers.
- Strawson, Peter F. (1992): *Analysis and Metaphysics: An Introduction to Philosophy*, Oxford: Oxford University Press.
- Williams, Bernard (1980): “Internal and External Reasons”, reprinted in: Williams, Bernard (1981): *Moral Luck*, Cambridge: Cambridge University Press.
- Woodfield, Andrew (1976): *Teleology*, London: Cambridge University Press.

