Part I

Action Theory

1

Action Explanation

RALF STOECKER

1. Introduction

Understanding action explanation is not only an interesting philosophical task per se, it is of utmost importance for action theory in general, since according to most philosophers, it belongs to the very essence of actions to be explainable in specific ways. Action explanation is also at the core of Davidson's theory of action. "Actions, Reasons, and Causes" (ARC), his first and most prominent contribution to action theory, was explicitly devoted to the 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?" (*Essays on Actions and Events* (EAE) 3). And he further developed his view on action explanation in many of his later articles. Moreover, for Davidson, the whole enterprise of interpretation, that is, ascribing intentional attitudes as well as linguistic meaning to our fellow people, aims at the explanation of actions (as part of a "unified theory of meaning and action," *Problems of Rationality* (POR) Ch. 10). In this sense, action explanation forms the root of Davidson's philosophical work in general.

Davidson's account of action explanation has also been one of the most influential parts of his philosophical work. Under the heading of the "causal theory of action," it evolved into a standard conception of agency that is not only taken for granted by many action theorists, but also by philosophers of mind and ethicists. And when the standard theory comes under attack, critics still refer to Davidson.

Despite its prominence, however, there has been comparatively little scholarly effort invested into a precise interpretation of Davidson's general account of action explanation. Apparently, the overwhelming success of ARC and the publication of a selection of his earlier articles under the title *Essays on Actions and Events* (EAE) in 1980 misled philosophers working in action theory into believing that they could confine themselves to the critical study of these papers. However, in order to understand Davidson's theory of action, and in particular his theory of action explanation, it is important to take into consideration Davidson's writings on radical interpretation and intentional attitudes as

A Companion to Donald Davidson, First Edition. Edited by Ernie Lepore and Kirk Ludwig.

well (many of which are to be found in the second collection of his earlier articles: *Inquiries into Truth and Interpretation* (ITI)), and also to realize that Davidson developed and considerably refined his views in later articles (which nowadays are easily accessible in the three additional volumes of Davidson's Collected Essays: *Problems of Rationality* (POR), *Subjective, Intersubejctive, Objective* (SIO), and *Truth, Language, and History* (TLH)), as well as in numerous replies to criticisms. As it turns out, the overall picture, resulting from a comprehensive study of Davidson's oeuvre, is in many respects quite different from the standard causal theory of action.

2. Actions and Their Rationalization

Actions, according to Davidson and most other philosophers of action, are events. Events in turn are spatiotemporally individuated entities, standing in part-whole relationships, as well as in causal relations to other events (cf. EAE Essays 8–10 and Appendix B). For Davidson, however, only events of a special kind can be actions, namely bodily movements (EAE 59, POR 101–106). This claim emerges from his account of a prominent feature in the ascription of agency: the employment of the "by"-locution. When a person acts, she typically does various things *by* doing others, for example, she illuminates a room *by* turning on the light, which in turn she does *by* flicking the switch (EAE 53). The observation that ascriptions of agency often involve "by"-sentences gave rise to the question of the relationship between the actions that are ascribed on both sides of the "by," which in turn fuelled a lively debate in the early days of modern action theory. Davidson's answer to this question is radically "coarse grained." In saying that the agent illuminates the room by turning on the light, which she does by flicking the switch, one is picking out one and the same bodily behavior: a finger movement touching the switch. The agent's actions of illuminating the room, turning on the light, and flicking the switch are identical. Yet if the actions involved on both sides of the "by"-locution are identical, then, given that ultimately people always act by moving their bodies, it follows that all actions are bodily movements. (See the entry on the individuation of action.)

Not all bodily movements are actions, however. For Davidson to call a bodily movement an action is to say that it can be explained in a specific way that he calls "rationalization." In this sense, the possibility of giving action explanations is essential for agency.

There are many different kinds of rationalizing explanations of actions. One can explain an action, for example, by citing an intention (*A man nails boards together with the intention of building a squirrel house*, EAE 83), an expected outcome (*I pour you a shot because it will sooth your nerves*, EAE 8), or the object of a want (*I went into the store because I want that gold watch in the window*, EAE 6), by giving new descriptions of the same action (*"Why are you bobbing around that way?" – "I'm knitting, weaving, exercising, sculling, cuddling, training fleas."* EAE 8), by using a "by"-sentence (*By setting fire to the bedding, Smith burned down his house*, EAE 47), or by pointing out the aim of the action (*Smith burned down the house in order to collect the insurance*, EAE 47).

It is therefore not quite correct to object that Davidson only considers a special kind of sophisticated action explanation (Thompson 2008: 86); he is well aware of the whole variety of everyday explanations. Yet for Davidson, all these explanations share a common core: they explain an action by leading us "to see something the agent saw, or thought he saw, in his action" (EAE 3). As a minimum, they describe the agent as having a favorable attitude (a "pro attitude") to actions of a certain kind (e.g., building a squirrel house), plus the belief that the action to be explained (e.g., the action of nailing the boards together) is of the respective kind (EAE 3–4). In ARC, Davidson calls such a pair of attitudes a "primary reason." Explanations that merely point out the primary reason usually are not very interesting. According to Davidson, however, the more interesting explanations, which explain the action without explicitly mentioning the primary reason, still indicate a primary reason and thereby show what the agent saw in the action. ("'I want that gold watch in the window' is not a primary reason and explains why I went into the store only because it suggests a primary reason – for example, that I wanted to buy the watch" (EAE 6).)[1]

Michael Smith has called Davidson's claim that there has to be a primary reason for every action the "Humean dogma" (Smith 1994: 92). It is not clear, however, whether recent debates on Humean (or internalist) versus Kantian (externalist) theories of motivation really are of concern for Davidson's account, because his understanding of pro attitudes is presumably much broader than a typical "Humean" understanding of desires. What seems to be clear, though, is that Davidson's view is challenged by authors who deny that there have to be primary reasons at all for actions, like Rosalind Hursthouse in her conception of arational actions (Hursthouse 1991), or philosophers who doubt that reasons are mental attitudes (e.g., Alvarez 2010; Bittner 2001; Dancy 2000; Hacker 2007; Stoutland 2001).

In any case, rationalizing explanations do more than merely point out that an agent has a certain favorable attitude toward one of her actions. They express that the agent acts *on* her attitude. Action explanations hence have two elements: in addition to expressing what the agent sees in the action, they claim that what she sees is a reason *for which* she acts. This additional element of action explanations is not redundant since, first, there are situations where an action is attractive to the agent in a number of different respects, yet only one of them is responsible for her doing what she does, and, second, there are situations where a certain behavior fits the agent's demands perfectly, and still the behavior is merely a reflex or induced from the outside and hence no action at all. In both kinds of situation, pointing out why it seems favorable for the agent to do what she does is not sufficient for explaining it.

The interesting question is how to understand this additional element in action explanations. Following Aristotle, one might be tempted to think that the action has to be somehow set in motion by the agent, for example, by an act of will or volition of hers. When Davidson wrote ARC, however, there was almost unanimity in action theory that rationalizations should not be construed as causal explanations, and in fact even that they *could* not be construed in this way. Proponents of this view, very much influenced by Ludwig Wittgenstein, were, for example, G.E.M. Anscombe (Anscombe 1957) and A.I. Melden (Melden 1961). It was Davidson's express intention in writing ARC to combat these views and to show that a causal understanding of the additional element in action explanations is not only feasible but highly plausible, and he devoted a considerable part of this philosophical work to elaborating and refining this causal position. Overall, Davidson's argument combines three lines of attack: (i) a move that is occasionally called "Davidson's challenge" (cf. Mele 2000: 279); (ii) a refutation of attempts that were supposed to show that rationalizing explanations cannot be causal; and (iii) a positive account of the causal character of rationalizing explanations.

3. Davidson's Challenge and the Problem of Wayward Causal Chains

Davidson repeatedly made clear that in fact, he was much in agreement with most of what his Wittgensteinian colleagues said about the way rationalizing explanations fit actions into a pattern of the agent's behavior (e.g., ITI 159). What is missing in the Wittgensteinian proposals, however, is an account of *how* one could explain the occurrence of an action by merely showing that it fits into a certain pattern (EAE 10). According to Davidson, this is why the explanations should be regarded as causal. We know that causal explanations provide a possible way of explaining occurrences; it is therefore a good first guess to assume that action explanations are causal, too. The challenge he puts to the anti-causalist is to put forward an alternative account of the explanatory force of action explanations that does not make use of causality.

Action theorists have repeatedly taken up Davidson's challenge and in particular have tried to strengthen teleological readings of rationalizing explanations instead (e.g., George Wilson, and more recently G.F. Schueler, Scott Sehon, and Timothy O'Connor; cf. Sehon 2010). This is not the place to assess these proposals, which have gained much popularity during the last couple of years, yet it is necessary to mention that they are frequently motivated by skeptical doubts about whether Davidson himself could meet his challenge, that is, whether Davidson and the causalists are in fact better off with respect to Davidson's challenge than their opponents. This skepticism is based on the assumption that Davidson can only claim to have given an adequate account of the extra element in action explanations if his characterization of these explanations sufficed to distinguish actions from nonactions. If actions are by essence explainable in a specific way, and if it is characteristic of this specific way that the explanation expresses, first, what the agent saw in the action and, second, that the action's taking place was caused by the agent's reasons for it, as Davidson apparently maintains, then there should be no instances of bodily behavior that satisfy these two conditions and yet are not actions. But in fact, there are such cases, that is, cases of so-called *wayward causal chains*, where a person's behavior is favorable from her own perspective and is, moreover, caused by her pro attitude, but the behavior still is not an action.

Davidson was one of the first action theorists to formulate this problem and he contributed several creative examples to the debate, for example, the following: "Thus a man might want to break a pot, and believe that by stamping on the floor he will cause the pot to break. The belief and desire cause him to stamp, but the stamping has no direct effect on the pot. However, the noise makes a bystander utter an oath which so offends the agent that he swings around, accidentally knocking over and breaking the pot. The agent had a motive for breaking the pot, and the motive caused him to break the pot. But he had no motive in breaking the pot: it was an accident" (EAE 264). Although the destruction of the pot was caused by the appropriate pair of belief and pro attitude, it was not an action; hence a causal relationship between primary reason and behavior is not sufficient for agency.

Unfortunately, though, Davidson also did much to obscure his own position when he wrote in his seminal article "Freedom to Act," "What I despair of spelling out is the way in which attitudes must cause actions if they are to rationalize the action" (EAE 79, cf. also EAE 264, POR 106), without making it sufficiently clear that for him this was not a weakness of his approach but evidence for something important about action explanations. At least, defenders of a noncausal, teleological view of action explanations have taken Davidson's confession as a strong confirmation of their view that causal accounts of action explanations cannot handle the phenomenon of wayward causal chains and, hence, do not meet Davidson's challenge themselves.

4. The Logical Connection Argument

Given the initial plausibility of causalism and the strength of Davidson's challenge, it is worthwhile to turn to the second line of Davidson's vindication of causalism, the question whether there are some principled obstacles that speak against the very possibility of a causal account of rationalizing explanations.

In the early 1960s, probably the most prominent objection against a causal understanding of action explanations was the so-called *logical connection argument*. The argument is based on the assumption that causal explanations derive their explanatory power from a causal relation between the event to be explained and another event by which it is caused. The occurrence of such a sequence of two distinct events is an empirical fact. Hence, it is assumed any causal explanation of an event with reference to another causally efficacious event must be expressed in a contingent sentence. Rationalizations, however, were said to be based on a kind of logical connection between the agent's attitudes and her action, since her attitudes are only explanatory if their content somehow fits the action. Rationalizations therefore could not be causal explanations.

Despite its prominence in the early days of action theory, the logical connection argument has at least two weak points, both of which Davidson tackles in ARC. First, the presumed logical connection between reasons and actions is extremely shaky. One simply cannot infer from the information that a person has a pro attitude toward illuminating the room and that he knows that flicking the switch would illuminate the room to the conclusion that the person will illuminate the room (perhaps he is too lazy to get up or too much occupied with building a squirrel house). Second, the whole idea of a mismatch between causal relations and logical relations rests on a category mistake. Whether a sentence expresses a causal relationship depends on *what its expressions refer to*, while whether it is logically true or contingently true depends on *which expressions are used*. The sentence "The event of United Airlines Flight 175 crashing into the southern tower of the WTC caused the collapse of the building" is true, and it is a contingent truth; while the sentence "The event that caused the collapse of the southern tower of the WTC caused the building's collapse," although true as well, is almost a logical truth (cf. EAE 13–15). Given the number of impressive philosophers who have once defended the logical connection argument in one form or another, one may wonder whether the argument really was based merely on a sort of logical blunder (cf. Kenny 1975: 117–120), but in any case, it almost disappeared from action theory after Davidson's devastating criticism.

Much stronger than the logical connection argument are objections that point to apparent tensions between Davidson's causal understanding of rationalizations and other parts of his philosophical work. Particularly conspicuous are two seeming conflicts with Davidson's understanding of causation. First, action explanations seem to provide the wrong kind of explanantia for causal explanations, and, second, there seems to be a contradiction between Davidson's view that causal relations entail strict laws covering the succession of cause and effect and his insistence that there are no (and could be no) laws that cover ascriptions of mental attitudes and ascriptions of actions.

5. Reasons as Causes?

In the literature, Davidson's causalism has been frequently summed up with the catchphrase: *reasons are causes*, and Davidson himself has occasionally used similar phrases (e.g., EAE 4, 233). These bold slogans, however, apparently contradict Davidson's claim in his articles on causation that the relata of causal relations are events (EAE ch. 7). For Davidson beliefs, desires and other attitudes are not events and consequently are not even of the right kind to be causes of actions. So, how could he still maintain that reason explanations are causal explanations?!

Davidson had already anticipated this objection in ARC (12–13). But since many critics take it that his response is basically to be found in the following sentence, "States and dispositions are not events, but the onslaught of the state or disposition is" (EAE 12), they were understandably not convinced. The explanatory value, for example, of a desire to travel to Australia someday is obviously not captured by the fact that it was once acquired, years ago. Hence, Davidson's response apparently fails to dispel the objection.

There is another objection that points into a similar direction but was, to my knowledge, never addressed by Davidson. Harry Frankfurt argues that action explanations cannot be causal explanations because reasons only explain actions if they are simultaneous with the action, while causes and effects occur in temporal succession (Frankfurt 1988). That a man wants to illuminate the room explains his flipping the switch only if, at the moment of his moving the finger, he still wants to illuminate the room, while a causally efficacious desire might well have ceased when the effect takes place. Consequently, according to Frankfurt, reasons are not causes.

6. The Role of Laws in Action Explanations and the Causal Relevance of Mental Properties

The second objection that points to a tension between Davidson's conception of causation and his causal understanding of action explanations concerns the role of laws. The objection can be rather different in kind. The most radical one is at the same time the one which is easiest to refute. It starts from two premises: first, every causal relationship is covered by a strict law, and second, it is impossible to formulate strict psychophysical laws. It then draws the conclusion that there could be no causal relations between mental events and physical events (e.g., actions).

The argument is not conclusive, though. As Davidson has shown in his famous argument for anomalous monism, there is another, alternative conclusion that can be drawn, namely that all (causally connected) mental events are physical events. For Davidson, every mental event is at the same time a physical event, that is, an event to which one can refer in purely physical language. Consequently, it is perfectly possible that mental events (described physically) fall under strict covering laws, although there are no psychophysical laws, and hence it is possible that mental events are causes of physical events. Davidson's presumptions that every causal relationship is covered by a strict law and that there are no strict psychophysical laws do not per se contradict his causal understanding of action explanations. (See the entries on causation and anomalous monism.)

Although this response is formally impeccable, to many philosophers in action theory and the philosophy of mind, it seemed insufficient and beside the point and therefore led to an extended debate on agency and mental causation in the late 1980s and early 1990s (cf. the articles in a special issue of *Philosophical Perspectives* 3 (1989) and in Heil and Mele (1993)). What these authors complained about was that in order to defend the causal understanding of action explanations, it is not enough to show that mental events could be causes of physical events. What one has to prove is that they can cause something *qua* being mental, or to put it slightly different: one has to prove that mental properties could be *causally relevant*, that they are not merely epiphenomenal. Fred Dretske has illustrated the objection by a telling analogy: "Meaningful sounds, if they occur at the right pitch and amplitude, can shatter glass, but the fact that these sounds have a meaning is surely irrelevant to their having this effect" (Dretske 1989: 1). Hence, what defenders of causalism have to show is that the property of being mental is not as causally inert in action explanations as the libretto of the soprano is who shatters glass with her voice.

The objection can take different lines. Jaegwon Kim has proposed it in an ontologically rich sense as an argument about the causal role of properties. For Kim it is part of our physicalistic understanding of the world that every event can be fully explained with recourse to preceding events and their physical properties. Accordingly there is no causal role for mental properties, unless they are in fact physical properties, which leads to his criticism of so-called token identity theories (like Davidson's) and to his defense of type identity theories. From Davidson's perspective, however, this objection is misguided right from the beginning. Since properties are strictly speaking not part of his ontology at all, he need not worry about their causal relevance or inertness. Moreover, as he frequently emphasizes, he rejects the whole idea of an event causing another *qua* having a certain property or *qua* being an exemplification of a property (TLH 188).

The objection is much stronger yet, if it is not understood ontologically but as an argument about the *explanatory power of reason explanations*. According to this reading, Dretske's example shows that we learn very little about *why* the glass broke when we are merely told that it fell into pieces because a woman declared her passionate love (unless we are also informed that she is a soprano singer and that her oath of love was at an extremely high pitch). The way the cause is described, it seems, is not useful in explaining why the glass broke. Another, quasi opposite, example for explanatory feebleness has already been mentioned in the discussion of the logical connection argument. It is true that the event that caused the collapse of the southern tower of the

WTC caused the building's collapse, but as being almost trivially true it has no explanatory value at all. What these examples show is that one has to say much more about the explanatory role of rationalizing explanations than merely that they are causal explanations. One still has to explain *how* they could be causally explanatory. Insofar as the causal theory has failed to do so, again, Davidson's challenge has not been met (cf. Follesdal 1985: 315).

This is where finally the objection seems to get grip that there is a tension between Davidson's causal understanding of action explanation and his conception of the relationship between causation and laws. For it is tempting to go back to the laws that cover causal relations in order to account for the explanatory value of causal explanations. Given that they are strict in the sense of being universally true without exception, these laws seem to provide ideal causal explanations because they allow an inference from the occurrence of the cause to the occurrence of the effect; they provide, in the words of C.G. Hempel, a kind of deductive-nomological explanation (Hempel 1965: 347 ff.; cf. EAE ch. 14). For Davidson, however, who denies the very possibility of strict psychophysical laws, rationalizations could not be supported by covering laws, and therefore could not be explanatory in the sense mentioned. There seems to be an inevitable dilemma that one must either give up the idea that action explanations are causal explanations or agree that the mental explanantia fit into strict laws.

In should be noted, however, that there is further difficulty for the covering law account of the explanatory force of action explanations, which is independent of the anomalousness of the mental. Typical action explanations which, for example, refer to a desire of the agent usually do not look like instances of regularities, since: "Far more often than not people fail to perform any action at all to achieve a desired end, even though they believe or know the means are at hand; and no one ever performs all the actions he believes will lead to the end. [. . .] If we were to guess at the frequency with which people perform actions for which they have reasons (not necessarily adequate or good reasons, but reasons in the simple sense under consideration), I think it would be vanishingly small" (EAE 263–264). Hence, according to Davidson, if action explanations were based on regularities between what a person wants and what she does, most of our everyday explanation would be almost worthless. Neither would it help to try to improve these regularities by making them more specific and case sensitive (cf. Lanz 1993). Hence, whether there are any psychophysical laws or not, the explanatory force of action explanations apparently is not due to being an instance of a deductive-nomological explanation. But then, how can it be accounted for instead?

This is the point to switch from the second line of Davidson's defense of causalism, that is, his attempts to refute objections that try to show that causalism is untenable for him, to the third line, namely his positive account of causal explanations in general and reason explanations in particular. (See the entry on causation.)

7. Singular Causal Statements and Causal Explanations

There is one conspicuous feature that many of the objections against Davidson's causalism share. They seem to go much too far because they apparently prove that the bulk of common sense explanations (and perhaps also many explanations in science) are not causal, either. Davidson had already made this point in ARC and restated it again in later articles. First, with respect to the objection that reasons could not be causes because they are not events, Davidson points out that the same is true of many ordinary causal explanations: "the bridge collapsed because of a structural defect; the plane crashed on takeoff because the air temperature was abnormally high; the plate broke because it had a crack" (EAE 12). Neither structural defects nor temperature or cracks in the dishes are events, yet they allow for causal explanations of events like the collapse of a bridge, the crash of a plane or the breaking of a plate. Second, with respect to the objection that action explanations do not allow for covering laws, he writes: "I am certain the window broke because it was struck by a rock – I saw it all happen; but I am not (is anyone?) in command of laws on the basis of which I can predict what blows will break which windows" (EAE 16). Since none of our everyday concepts presumably will turn up in the strict laws that cover causal relations, in this respect, action explanations fare no worse than other everyday causal explanations.

This is where Davidson's positive account of causal explanations sets in, which is based on two crucial distinctions. First he distinguishes two different understandings of the expression "cause" (EAE 161–162; cf. Davidson 1993: 288, 1999: 638). Either the expression is used as a two-place predicate to be filled in by names of events ("United Airlines Flight 175's crashing into the southern tower of the World Trade Center," "the collapse of the building") or it is used as a connective between sentences, for example, between "the plate broke" and "the plate had a crack." In the first, "strict," sense, "cause" occurs in singular causal statements, while in the second, "broad," sense, it is used for causal explanations. Obviously, Davidson's claim that reasons are causes has to be understood in the latter sense.

However, even if one has to distinguish single causal statements from causal explanations, they have to be connected. Causal explanations must be somehow dependent on causal relations. And if action explanations are causal explanations, they must be based on causal relations, too. The question is, how?

8. Strict Laws, Generalizations, and Causal Concepts

At this point, Davidson's second crucial distinction comes into play. One has to distinguish between strict laws of nature that are entailed by causal relations and the generalizations on which familiar causal explanations are based (PoR 113). While the existence of laws of the former kind may very well be unknown to anybody, generalizations of the latter kind clearly play an important role in causal explanations. In *Laws and Cause* (TLH, ch. 14). Davidson sketches an almost Kantian picture of events as being individuated according to the demands of our explanatory interest in assigning regularities to the world: "It is not surprising, then, that singular causal statements imply the existence of covering laws: events are changes that explain and require such explanations. This is not an empirical fact: nature doesn't care what we call a change, so we decide what to count as a change on the basis on what we want to explain, and what we think available as an explanation" (TLH 212). The principle of the nomological character of causality, according to which every causal relation entails a strict covering law, is an *a priori* truth. It is an essential part of our metaphysics of events (cf. also Davidson 1985: 227).

But as Davidson already made clear in his famous argument for anomalous monism, "it is possible (and typical) to know of the singular causal relation without knowing the law or the relevant descriptions" (EAE 224). This is what our nonstrict everyday generalizations are good for: "Knowledge requires reasons, but these are available in the form of rough heteronomic generalizations, which are lawlike in that instances make it reasonable to expect other instances to follow suit without being lawlike in the sense of being indefinitely refinable" (EAE 224). It is characteristic for our everyday causal wisdom and also for most scientific knowledge that we base it on generalizations that we believe can, in principle and by reference to the basic laws of nature, be shown to be trustworthy under the specific circumstances where we employ them – although in fact we are not and perhaps nobody ever will be able to do it. And we express our conviction that we are faced with a case where they are trustworthy by using the language of causality. In this sense, claiming a causal relationship always is an unfulfilled promise, "a cloak for ignorance" (EAE 80), that deep down in the causal fabric of the world, the succession of events under question can be shown to be an instance of a strict law of nature. And our familiar rules and regularities provide the evidential base for such claims. We feel confident to say that the short-circuit caused the fire (EAE 151) because we have pretty good general knowledge of household electricity and its dangers, which we believe would not evaporate if some super physicists inquire more deeply into the relationship between the particular short-circuit and the particular fire under question. In this sense, we make the occurrence of an event (the fire) intelligible with reference to another, earlier event (the short-circuit), since we make the justified (although certainly not infallible) claim that the one had to follow the other due to the laws of nature.

Therefore, the fact that there are no strict laws that cover action explanations is no reason at all to deny that they are causal explanations; on the contrary, it is something that is shared by (almost) all causal explanations: "It is often thought that scientific explanations are causal, while explanations of actions and mental affairs are not. I think almost exactly the reverse is the case: ordinary explanations of action, perception, memory, and reasoning, as well as the attribution of thoughts, intentions, and desires, is riddled with causal concepts; whereas it is a sign of progress in a science that it rids itself of causal concepts" (PoR 96).

There remains the other problem, however, that there are not even rough and ready generalizations that support typical action explanations; in this respect, they are unlike, for example, the explanation of the fire by the short-circuit. But as Davidson repeatedly points out, many everyday causal explanations work differently, anyhow. They are based on *causal concepts*. The notion of "causal concept" plays a dominant, though frequently neglected, role in Davidson's work on agency and the mind, despite the fact that already in the introduction to his *Essays on Actions and Events*, he writes, "The theme [of the book] is the role of causal concepts in the description and explanation of human action" (EAE xi). In another article, he says: "In fact the causal character of the concepts used in talking about action is an essential part of what must be grasped in coming to a clear view of the nature of action explanation" (PoR 105). What Davidson means by "causal concepts," however, are not terms like "cause," "produce," and so on, but predicates that either imply a certain causal role (e.g., "having a sunburn," which entails that the bearer's burned skin was caused by being exposed to the sun (PoR 121)) or which imply certain causal powers or dispositions (e.g., "being biodegradable," which entails that under certain circumstances it will decompose by natural biological processes (PoR 95)).

Both kinds of causal concepts are important for understanding the explanatory force of action explanations: "it is part of the concept of an intentional action that it is caused and explained by beliefs and desires; it is part of the concept of a belief or a desire that it tends to cause, and so explain, actions of certain sorts" (SIO 217). As already mentioned at the beginning, according to Davidson, it is essential for an event being an action that it can be causally explained by reference to a primary reason, that is, a pair of a belief and a desire. "Action" is in this respect similar to "sunburn." And (which is a less familiar claim) it is essential for being a belief or a desire, or any propositional attitude whatsoever, that they are dispositions or causal powers of the agent. "As I already pointed out, beliefs and desires have causal powers, and that is why they explain actions" (PoR 112, cf. (Mayr 2011)). In order to understand Davidson's account of action explanation, one has to elucidate his view of causal powers.

9. Causal Powers

According to Davidson, a causal power is "a property of an object such that a change of a certain sort in the object causes an event of another sort" (EAE 64), where this second event could either be occurring within the same object (as in the causal power of being soluble) or somewhere else (as in being a solvent). For Davidson, causal powers play an important role in everyday causal explanations. They "encapsulate the relation between causality and laws" (TLH 214). Saying, for example, that certain wrappings are biodegradable is expressing our wisdom that under the right condition (e.g., a compost heap), it is probable that certain natural processes (having to do, e.g., with earthworms and bacteria) cause the decomposition of the wrappings. Consequently, we can give a causal explanation of the disappearance of a particular package in the heap after a few months by simply saying that it was biodegradable. It had the disposition to be caused to dissolve.

Such an explanation by reference to a disposition is far from being pointless. After all, the package could also have vanished because it was removed or burned. Although all causal concepts share the "cloak of ignorance" with respect to the specific underlying natural processes, they express a practical causal wisdom that justifies an explanatory claim. Much of what we know about the world is preserved either in generalizations, like the one that short-circuits tend to cause fire, or in our ascription of causal powers to things, for example, to paper wrappings, that they are biodegradable. They allow us to give causal explanations, although these explanations entail that there has been some natural process about which we have at best a very dim idea. The remaining uncertainty (the "cloak for ignorance") is expressed in the explicit reference to causality. Whenever we give a causal explanation, we express our confidence that a certain relationship obtains which we are unable to make explicit due to our lack of knowledge.

Given Davidson's account of causal concepts and particularly of causal powers, many of the objections against his causal understanding of agency dissolve. It is not a problem any longer that propositional attitudes are not events and hence cannot be causes, since causal explanations need not necessarily refer to any causally efficacious event. Sometimes they do (e.g., to a short-circuit), but occasionally they do not, for example, when they explain the vanishing of the paper wrapping with reference to its biodegradability. Therefore, beliefs and desires as causal powers can causally explain an action without being events. Of course, in order for the explanation to be true there have to be some causally efficacious events that cause the action, for example, the onslaught of a desire or a glimpse of the desired object, and so on. But frequently, we have no idea what the particular event was and still can easily explain the action by rationalizing it.

The account also furthers a better understanding of the phenomenon of "wayward causal chains," since it can be shown that the phenomenon is an instance of a general feature of explanations by dispositions and causal powers. Imagine that the paper wrapping has vanished from the compost heap not because it dissolved but because the owner picked it out as an illustrative sample for her children to teach them about biodegradability. In this case, its vanishing is still due to its being biodegradable, yet it is at least severely misleading to say: it has vanished because it is biodegradable, since the vanishing was not an *instantiation* or an actualization of the disposition of being biodegradable. In the same way, it can be argued that cases of wayward agency are not actualizations of the causal powers of the agents' intentions, although they are in some way due to them (Stoecker 2003). The fact that typical causal explanations display the same phenomenon of "wayward causal chains," far from falsifying Davidson's causal account, speaks very much in its favor.

We are left, however, with the question of what kind of causal power propositional attitudes are. As already mentioned, usually there is almost no regularity between having a certain desire and performing a particular action. In this respect propositional attitudes are very unlike disposition like biodegradability, fragility, solubility, and so on. What is still missing is an account of propositional attitudes as causal powers, which in the end leads to Davidson's theory of radical interpretation.

10. Propositional Attitudes as Causal Powers

One big difference between ascriptions of propositional attitudes on the one hand and of dispositions like being biodegradable on the other is that the former are relational. To say that an agents wants to build a squirrel house is to relate him to a "propositional content," namely that he builds a squirrel house. According to Davidson, the best way to construe this connection is as a special kind of linguistic maneuver. The speaker who attributes this desire to the agent says something like: "That's what he wants: He builds a squirrel house." The speaker attributes a desire to the agent pointing to something he, the speaker, utters, as a sort of specimen. Since the specimen sentence is not logically connected with the attribution sentence but only attached, Davidson calls his proposal a "paratactic account" of propositional attitude attributions (SIO 76–77, cf. ITI 106). In order to understand how propositional attitudes could be causal powers, though, one has to clarify the punch line of this linguistic practice. Why are we interested at all in relating an utterance of our own to another person?

The first step in Davidson's theory is to notice a surprising parallel to other dispositions ascribed relationally: for example, having a certain weight, length, temperature, and so on (EAE 220, SIO 59–60). That the water starts boiling in the kettle obviously is not due to its having some temperature (everything has), but rather due to its temperature being 100° C. In this way, one can explain the boiling of the water by relating the water to a number. Moreover, depending on the scale we use, assigning numbers can allow for complex dispositions (e.g., in a very simple example, we might explain the crash of an elevator with recourse to the culminated weight of the passengers, none of whom had the causal power to endanger the lift alone).

In the same way, according to Davidson, one can explain actions by relating agents to sentences or utterances (SIO 74–75). To say that the agent picks up the hammer because he wants to build a squirrel house is to explain his action of fetching the hammer as an actualization of a causal power that we keep track of ("measure") by relating the agent to the utterance "He builds a squirrel house." And as we can add up weights into more complex causal powers, we can also combine different propositional attitudes into more complex dispositions, for example, the desire to build a squirrel house and the belief that squirrel houses have to be fixed with screws instead of nails. Given both attitudes, it is no longer possible to explain the agent's act of fetching the hammer as by reference to the agent's disposition which we keep track of with by the sentence "He builds a squirrel house," while an act of getting a screw driver would still be perfectly explainable, and so on. The way these causal powers combine is provided by the basic principles of rationality.

It is still not clear, however, how this works as a causal explanation. At this point, Davidson's account of action explanations leads back to his theory of interpretation in the philosophy of language (cf. ITI Part 3). Understanding what people say, as a semantic enterprise, relates the speakers' utterances to the world. According to Davidson, our basic tool for relating speaker's utterances to the world is a *theory of truth* for the language of the speaker, which is even possible for alien speakers, that is, in a situation of what Davidson calls "radical interpretation." A theory of truth assigns truth conditions to the sentences uttered. It might say, for example, that the sentence "Schnee ist weiß," uttered by a speaker at a time, is true if and only if snow is white. As the logician Alfred Tarski has shown, it is possible to provide a finitely axiomatized truth theory for a language that contains quantifiers. Davidson and others have shown how to extend the sort of truth theory Tarski developed to natural languages that contain context sensitive expressions and types of sentences that do not occur in the formal languages that Tarski discussed. (For instance, Davidson's paratactic account was originally a suggestion to incorporate indirect speech into a truth theory.) Although it is difficult to formulate such a theory explicitly, Davidson assumes that: "All understanding of the speech of another involves radical interpretation" (ITI 125).

In radical interpretation, a theory of truth is treated as part of an empirical theory that, inter alia, ascribes to the speaker something like a representation of the world (ITI 133 ff., THL 35–36). The theory has to fulfill formal and empirical constraints (ITI 150 ff.). Empirical evidence for an adequate construction of the theory is found primarily in the vicinity of the speaker, where the interpreter can (provisionally) regard the speaker's utterances as being prompted by what is the case in her immediate surroundings (as in Quine's famous example of an alien speaker who utters "Gavagai" while a rabbit runs past). At least at these "edges" of the language, that is, indexical observation sentences, it must be granted that speech is causally embedded in the world (SIO 189).

Given that these empirical constraints are met, we can use our own sentences to "measure" (or keep track of) actual and even possible utterances of the speaker. This is where the concept of meaning is located in Davidson's philosophy of language (ITI, Essay 2), for example, when we say: the alien speaker's sentence "Gavagai" *means* that: a rabbit is running by. We use our sentence "A rabbit is running by" as a way of keeping track of the meaning of "Gavagai" in her language, based on an empirical theory of truth for that language. In this sense, the theory of truth provides a tool to measure the meaning of the sentences of a speaker's language.

Interpretation is more demanding, though. Its point is not merely to describe a language but to describe a speaker, by uncovering a propensity of hers to utter sentences under certain circumstances that would allow for inferences about unobserved and counterfactual situations (e.g., that she is or would be prompted by a passing rabbit to utter "Gavagai") (ITI 174). Davidson calls this propensity the "holding true" of sentences. That is what connects his theory of truth with his theory of action. "The interlocking of the theory of action with interpretation will emerge [. . .] if we ask how a method of interpretation is tested. In the end, the answer must be that it helps bring order into our understanding of behavior. But at an intermediate stage, we can see that the attitude of holding true or accepting as true, as directed towards sentences, must play a central role in giving form to a theory" (ITI 161). It is part of the ability of a competent speaker of a language that many of her real or even potential and counterfactual utterances can be explained by recourse to her holding true sentences of her language. And these attitudes of holding sentences true are in turn based on the agent having been confronted with the world in situations like the running by of a rabbit. In this sense, Davidson can claim: "We justifiably assume [e.g.] that a person who is now disposed to hold that 'is a dog' is true of dogs came by that disposition through experiences of dogs" (POR 85).

Now, since we can keep track of what the sentences of the speaker's language mean by recourse to sentences of ours that express the truth conditions of these sentences (in whatever sense this gives their "meaning"), we can likewise use them to keep track of hold true attitudes, for example, the attitude of holding true the sentence "Gavagai." The result is an attitude of the speaker we can "measure" by the following: a rabbit is running by. According to Davidson, such an attitude is called a belief. The alien speaker who utters "Gavagai" and holds true "Gavagai" believes that a rabbit is running by. Instead of describing a speaker by recourse to sentences being held true, we can now describe her with recourse to states of affairs that express the conditions under which these sentences are true. On the one hand, this is not an arbitrary scale, since it is based on the assumption that the speaker's holding these sentences true is at least partially due to being confronted with the conditions under which they are true (e.g., that a rabbit is running by). On the other hand, it allows for the important fact that speakers occasionally hold sentences true which are in fact not true. To say that the speaker believes, for example, that a rabbit is running by is to describe him as someone who is disposed to behave as if a rabbit is running by whether this really occurs or not. "The concept of belief thus stands ready to take up the slack between objective truth and the held true, and we come to understand it just in this connection" (ITI 170).

The construction of a theory of truth for a speaker enables the interpreter not only to ascribe to him a web of beliefs, however, but a whole system of attitudes that are identified with recourse to the interpreter's characterization of the truth conditions of sentences in the speaker's language. This is due to the fact that human agents are not merely representing the world but take an evaluative stand towards it. We do not confine ourselves to holding some sentences true, but we also *prefer* some sentences to be true and others to be false. Hence, the way we "measure" agents by ascribing truth conditions to her can take different dimensions or perspectives. This is what we express in ascribing different kinds of attitudes that stand for different dispositions, although they concur in their propositional content (e.g., believing, doubting, hoping, or being afraid that a rabbit runs by). ("Our sentences provide the only measure of the mental" (SIO 77).)

Ascriptions of such attitudes, in turn, are subject to further constraints in addition to the empirical evidence and formal demands of a truth theory. They have to conform to certain principles of rationality that relate them to the behavior, in particular to the nonlinguistic behavior, of the speaker. Perhaps the most basic of these principles is the claim from ARC that in order to explain the behavior of a person with recourse to the system of her attitudes, there has to be a primary reason for the behavior (rendering it an action). Then there is the more demanding principle that the action must coincide with what the agent regards as being best on the basis of all relevant reasons ("principle of continence") (EAE 41). And finally, there are even more demanding principles that come into play when the system of attitudes is further enhanced with a second layer of numeric measurement: subjective probabilities and the relative weights of preferences. This additional feature of the attribution of propositional attitudes is already part of our everyday practice of action explanations ("He built a squirrel house instead of repairing the doghouse because he liked squirrels more than dogs."), but according to Davidson, it also opens the way for constructing a psychologically more respectable unified theory of meaning and action (cf. POR 10) that unifies the tasks of creating a theory of truth and a theory of subjective probability in decision theory.

11. The Explanatory Value of Action Explanations

Davidson's account of intentional attitudes has radical ontological consequences that are not always sufficiently noticed in the literature. "In thinking and talking of the weights of physical objects we do not need to suppose there are such things as weights for objects to have. Similarly, in thinking and talking about the beliefs of people we need not suppose there are such entities as beliefs. Nor do we have to invent objects to serve as the "objects of belief" or what is before the mind, or in the brain. Such invention is unnecessary because the entities we mention to help specify a state of mind do not have to play any psychological or epistemological role for the person in that state, just as numbers play no physical role" (SIO 60). According to Davidson neither intentional attitudes nor their contents should be taken seriously from an ontological point of view. To ascribe an intentional attitude to an agent is like ascribing a weight to her, and the sentence we use to keep track of the attitude (its "content") is analogues to the number expressing how much she weights.

Actions consequently are not the causal effect of propositional attitudes. Still, action explanations are causal explanations because they are based on a highly sophisticated disposition of human language users that is characterized by an intertwined web of attitudes, each of which is measured at the truth conditions of the agent's language, which in turn is grounded in the agent's reactions to her immediate vicinity, that is, in a causal relationship that has shaped the agent. Actions are actualizations of this disposition of the agent. When actions are explained in this way, what we learn from the explanation is not what caused the action. Instead, we learn something about the agent, namely the special idiosyncratic shape of her disposition to behave rationally in the light of her propositional attitudes (EAE 274). Learning for example that she wants to illuminate the room adds a certain strain to her disposition that is not already entailed in her being rational and that other people do not necessarily share or expect. And saying that she is acting *because* she wants to illuminate the room is to say that her behavior is an actualization of this disposition.

Acknowledgements

I would like to thank Jens Kulenkampff, Kirk Ludwig and Thomas Spitzley for numerous helpful suggestions and criticisms of earlier versions of the text.

Note

[1] It is evidence for the extraordinary influence of ARC that its terminology is still frequently used in the literature although subsequently to ARC Davidson himself immediately gave up the notion of "primary reason" and also made use of his second coinage "pro attitude" only in two other papers ("Intending" and "Problems in the Explanation of Action").

References

Alvarez, M.A. (2010). *Kinds of Reasons: An Essay in the Philosophy of Action*. Oxford: Oxford University Press.

Anscombe, G.E.M. (1957). *Intention*. Oxford: Blackwell.

Bittner, R. (2001). *Doing Things for Reasons*. Oxford: Oxford University Press.

Dancy, J. (2000). *Practical Reality*. Oxford: Oxford University Press.

Davidson, D. (1985). Replies to essays I-IX. In *Essays on Davidson. Actions and Events*, B. Vermazen and M.B. Hintikka (eds). Oxford: Oxford UP.

———. (1993). Reply to Ralf Stoecker. In *Reflecting Davidson*, R. Stoecker (ed.). Berlin; New York: De Gruyter.

———. (1999). Reply to Jennifer Hornsby. In *The Philosophy of Donald Davidson*, L.E. Hahn (ed.). Chicago; LaSalle, IL: Open Court.

Dretske, F. (1989). Reasons and causes. *Philosophical Perspectives* 3:1–15.

Follesdal, D. (1985). Causation and explanation: a problem in Davidson's view on action and mind. In *Actions and Events*, E. Lepore and B.P. McLaughlin (eds). Oxford: Basil Blackwell.

Frankfurt, H.G. (1988). The problem of action. In *The Importance of What We Care About*, H.G. Frankfurt (ed.). Cambridge, UK: Cambridge University Press.

Hacker, P.M.S. (2007). *Human Nature: The Categorial Framework*. Oxford: Blackwell.

Heil, J. and Mele, A.R. (1993). *Mental Causation*. Oxford: Clarendon Press.

Hempel, C.G. (1965). *Aspects of Scientific Explanation, and Other Essays in the Philosophy of Science*. New York: Free Press; London: Collier-Macmillan.

Hursthouse, R. (1991). Arational actions. *The Journal of Philosophy* 88(2):57–68.

Kenny, A. (1975). *Will, Freedom and Power*. Oxford: Blackwell.

Lanz, P. (1993). The explanatory force of action explanations. In *Reflecting Davidson*, R. Stoecker (ed.). Berlin: W. de Gruyter.

Mayr, E. (2011). *Understanding Human Agency*. Oxford; New York: Oxford University Press.

Melden, A.I. (1961). *Free Action*. London; New York: Routledge and Kegan Paul.

Mele, A.R. (2000). Goal-directed action: teleological explanations, causal theories, and deviance. *Noûs* 34:279–300.

Sehon, S. (2010). Teleological explanation. In *A Companion to the Philosophy of Action*, T. O'Connor and C. Sandis (eds). Oxford: Blackwell.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell.

Stoecker, R. (2003). Climbers, pigs and wiggled ears: the problem of waywardness in action theory. In *Physicalism and Mental Causation: The Metaphysics of Mind and Action*, S. Walter and D. Heckmann (eds). Charlottesville, VA: Imprint Academic.

Stoutland, F. (2001). Responsive action and the belief-desire model. *Grazer Philosophische Studien* 61:83–106.

Thompson, M. (2008). *Life and Action: Elementary Structures of Practice and Practical Thought*. Cambridge, MA; London: Harvard University Press.