

# 5

## Intention

Luca Ferrero

### 1. The Initial Account

Davidson's account of intentional action develops in three stages. In Davidson (1963) and Davidson (1971), he discusses intentional action in relation to the explanation and the ontology of action. His earlier view equates acting intentionally with being caused to act by a pair of appropriately related mental states (a pro-attitude and an instrumental belief) and denies the existence of intentions as distinct mental states. In Davidson (1969), he offers a more complex account of practical deliberation in terms of various kinds of evaluative judgments. In Davidson (1978), he discusses intentions for future action and reverses his earlier position about the nonexistence of intentions as distinct and irreducible mental states. In his later account, he argues that intentions are all-out evaluative judgments.

The starting point of Davidson's investigation of intentional agency is the notion of "intentional action." On the basis of his view of intentional action, Davidson discusses various related notions, including the idea of acting "with the intention to" do something, "intention" as a distinct mental state, an "intending to perform a future action."

Davidson (1971) is primarily concerned with the relation between the justification of action in terms of reasons and the explanation of action in terms of causes. Davidson argues that reasons both justify and cause intentional actions. Causation by reasons accounts for the intentional character of actions. More precisely, a particular action  $\phi$  is intentional if and only if it is performed for a reason, in the sense that it is caused (in the right way) by the combination of (i) a pro-attitude for actions of kind  $K$ , and (ii) a belief that that *particular* action  $\phi$  (under a description  $d$ ) is of kind  $K$ . This combination is what Davidson calls a *primary reason*. The primary reason both explains and rationalizes the action  $\phi$ . For instance, it is getting dark and I have a desire to illuminate the room. My desire is for actions of kind  $K$ : actions that result in the illumination of the room. I also have a belief that by flipping this particular switch, the room will be illuminated. The combination of this belief and desire provides the primary reason for my particular action  $\phi$  of flipping this switch, an action that belongs to the desired kind  $K$ , that of all the actions that result in the illuminating of the room.

To have a pro-attitude is to be favorably disposed toward actions of a certain kind. The particular action is only referred to in the cognitive component of the primary reason, in the instrumental belief that indicates that the particular action  $\phi$  is an instance of the desired kind  $K$ . Although Davidson often refers to pro-attitudes in general as "desires" or "wants" and to the way in which actions look favorable as their "desirability," pro-attitudes encompass a variety of conative attitudes, in a spectrum that goes from an urge to a sense of obligation or duty (Davidson 1971: 4, 6).

In doing  $\phi$  intentionally, the agent acts “under a description.” The agent cannot be said to be acting intentionally with respect to *all* the features of the action. Consider this case: “I flip the switch, turn on the light, and illuminate the room. Unbeknownst to me I also alert a prowler to the fact that I am home” (Davidson 1971: 4). According to Davidson, in this scenario, the agent performs only one action, the finger movement required for flipping the switch. This action can be described in many different ways, including in terms of any of its causal effects. Among these effects, there are the turning on of the light, the illuminating of the room, and the alerting of the prowler. The action can be described in any of these terms but it is intentional only under some of them. The action might be intentional under the description “flipping the switch” but unintentional under the description “alerting the prowler.” Attributions of intentionality are semantically opaque: even if the description “flipping the switch” picks out the same action as the description “alerting the prowler,” one cannot validly infer from *As* flipped the switch intentionally to *A*’s alerted the prowler *intentionally*.

The descriptions under which an action is intentional are determined by its primary reason (or primary reasons if there are more than one). More precisely, they are determined by the features under which the agent subsumes the action in the belief component (Davidson 1971: 5, 1978: 87). For instance, consider an agent who desires to illuminate the room and believes that action  $\phi$  – *as a particular flipping of the switch* – is expected to bring about the illuminating of the room. In this case, the action is intentional under the description “flipping the switch.”

Although Davidson does not explicitly state so, it also seems that the agent can be said to act intentionally under the description of the property toward which the agent has the pro-attitude (at least when the action succeeds in having that property). In the example earlier, the agent is acting intentionally also under the description “illuminating the room,” given that she flips the switch in light of her belief that this flipping is going to cause the satisfaction of her pro-attitude in favor of the illumination of the room.

By contrast, the prowler becoming alert to the agent’s presence in the room is a causal effect of the flipping no conception of which plays a role in the primary reason and which is unknown to the agent. As such, the alerting is not an aspect under which her action is intentional. Nonetheless, the alerting of the prowler is still something that the agent does, albeit unintentionally.

Davidson’s discussion of acting under a description captures a basic feature of intentional agency. But it does not explicitly address all of the possible ways in which the descriptions of an action might bear on its intentionality.

For instance, what should we say about the causal effects (or other properties of the action) that the agent foresees but do not feature in her primary reason? Let us imagine that the primary reason is only in favor of turning on the light. Say, the agent is an electrician who only cares about demonstrating that the electric circuit is working properly but does not care about illuminating the room. In order to test the circuit, it would be sufficient to turn on a very dim LED light that is unable to illuminate the room. But given that the bulb currently attached to the circuit is strong enough to illuminate the room, the electrician *predicts* that the room will be illuminated as a result of her flipping the switch and turning on of the light. In this case, both the flipping of the switch and the turning on of the light are intentional. But what about the illuminating of the room? Is it just an unintentional deed? And if so, is it unintentional in the same way as the *unforeseen* alerting of the prowler would be?

Another potential problem might arise for the description of an action that fails to achieve the intended goal. If I flip the switch in order to illuminate the room, but the electric circuit fails, am

I acting intentionally under the description “illuminating the room”? Could an action be intentional under a description that turns out to be false of it? This seems problematic given that “A intentionally  $\phi$ -ed” implies that “A  $\phi$ -ed,” and thus if A did not  $\phi$ , then A did not intentionally  $\phi$  (see Davidson 1963 fn 3).

It is not uncommon for intentional actions to fail at attaining the property *K* mentioned in the pro-attitude. These actions might still be intentional under the description that individuates them in the belief component. Consider the agent who desires to illuminate the room and believes that flipping the switch will bring about the illuminating of the room. If the flipping fails to illuminate the room (say, because the bulb is burnt out), the action can still be said to be intentional under the description “flipping the switch.” But what should one say if the action fails even at satisfying the description “flipping the switch”? Imagine that the switch is stuck and the agent is unable to flip it. Is there something that the agent does intentionally even if none of the descriptions mentioned in the primary reason are true of her action?

It might be plausibly suggested that what is intentional is the agent’s finger movement. Unlike descriptions such as “illuminating the room” or “flipping the switch,” it seems that the description “moving the finger” always apply to the action in question. This is so if moving the finger is a *primitive action*, that is, something that one does not do *by* doing something else (Davidson 1971: 59). Usually a primitive action is a bodily movement. An action that is not primitive is nothing other than a primitive action under a different description: a primitive action described in terms of some of its causal effects. (For further discussion of primitive actions, see Lepore and Ludwig (2013: ch. 3).)

Davidson’s account of the relation between intentionality, action descriptions, and primary reason might thus be modified so that the primary reason necessarily include or entail a reference to the description of the action in its primitive form. In this way, as long as the action is performed, there is always something that the agent does intentionally (in our example, moving the finger), even if the action fails either at accomplishing what the agent desires (illuminating the room) or at satisfying the nonprimitive description under which it was undertaken (flipping the switch).

A further complication arises because of Davidson’s statements about the relationship between acting intentionally and knowing what one is doing.

Davidson claims that one might be doing something intentionally and yet not know that one is doing it: “A man may be making ten carbon copies as he writes, and this may be intentional; yet he may not know that he is; all he knows is that he is trying” (Davidson 1971: 50, see also 60). But then he goes on to claim that: “Action does require that what the agent does is intentional under some description, and this in turn requires, I think, that what the agent does is known to him under some description” (Davidson 1971: 50).

Is the description under which one knows what one is doing the same as one of the descriptions under which the action is intentional? If so, the description would have to be a feature of the action that is guaranteed to be cognitively accessible by the agent and true of the action regardless of the action’s ultimate success. As before, a description of this sort might thus be the one under which the action is primitive (Davidson does not explicitly endorse this claim, although he seems to come close to it at Davidson (1971: 50).)

The notion of a primitive action might thus be crucial to tie together Davidson’s remarks about the relationship between the intentionality of an action with, on the one hand, the descriptions of the action and its primary reason(s), and with, on the other hand, the extent of the agent’s

knowledge of what she is doing. However, Davidson never explicitly clarifies how the descriptions of actions in their primitive guise both (a) fit into his account of the content of primary reasons and (b) relate to descriptions of actions in nonprimitive terms (which appear to be the descriptions under which many actions are usually said to be intentional).

The primary reason rationalizes the action in virtue of the logical relation between the contents of its components: the belief and the pro-attitude could be made into the premises of a practical syllogism from which one can deductively infer that the action has some desirable feature (Davidson 1963: 9). For instance, from the major premise “Any act of mine that results in the illuminating of the room has a desirable characteristic” and the minor premise “This action  $\phi$  will result in the illuminating of the room,” one can deduce the conclusion “This action  $\phi$  has a desirable characteristic.” A practical syllogism of this kind is part of the analysis of the concept of the reason with which one acts. It is not part of the agent’s practical reasoning, the reasoning about what to do. Davidson criticizes both Aristotle and Anscombe for taking the practical syllogism as a piece of practical reasoning, hence taking its conclusion to be that the action does not simply have a desirable characteristic but it is desirable, that is, worth doing (Davidson 1963: fn. 4; Davidson 1978: 98); whereas practical reasoning for Davidson has a more complicated structure, which involves the weighing of reasons (Davidson 1963: 16, 1969).

Davidson’s (1963) main goal is to argue that reasons are causes: that they not only rationalize the action but also explain it by causing it. To explain the actual performance of  $\phi$ , it is not sufficient to have a pro-attitude and a belief that, taken together, would make the action  $\phi$  intelligible. An agent might have several primary reasons that could be separately sufficient to rationalize the action. Explanation, however, requires that the belief/desire pair be the reason *for* which the agent did  $\phi$ . Davidson argues that this can be so only if that particular primary reason, rather than any of the alternative ones, *causes* the action. (For further discussion of this argument, see Lepore and Ludwig (2013: ch. 1).)

Hence, the sense in which a reason is “primary” is a matter of its contribution to the *explanation* rather than the justification of the action. This is not to deny that the rationalization provided by the primary reason offers some justification for the action, but it does so only in the “anemic” sense that it shows that “from the agent’s point of view there was, when he acted, something to be said for the action” (Davidson 1963: 9). This is because, in the order of justification, the existence of the primary reason only tells us that the agent looks favorably on the action given her belief that that action satisfies one of her pro-attitudes. More interesting questions about justification concern the specific pro-attitude in support of the action (including which properties of the action the agent looks favorably on and the kind of attitude she takes toward it – A desire? A sense of obligation? An urge? – and how these pro-attitude is to be weighed with respect to any competing considerations). The fact that a reason is primary is no indication of its strength in justification over competing reasons in support of alternative courses of action. It only shows that the primary reason is the particular belief/pro-attitude pair that was effective in causing the action among the agent’s other pairs of attitudes that provided actual alternative rationalizations (and thus potentially competing causal explanations) for that same action.

Notice that, as mental *states*, the pro-attitude and the belief cannot cause the action. They rather operate as causal conditions that can be triggered by the circumstances. It is only the “onslaughts” of the attitudes, which are actual *events*, that can play the proper role in the causation of action (Davidson 1963: 12).

Although reasons for action are causes, the causal relation between the attitudes and the action is neither identical nor necessarily isomorphic to their logical relation. Reasons are not causes in

virtue of the features that make them justificatory. According to Davidson, the laws of rational psychology are not laws of efficient causality. (This is a very important theme in Davidson's philosophy that will become more prominent in his defense of "anomalous monism" – see discussion in Lepore and Ludwig (2013: ch. 26).)

Beliefs and desires rationalize an action only if they cause it *in the right way*. The action is not intentional if it is caused by the attitudes in a *causally deviant way*.

A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally. (Davidson 1973: 79)

Davidson despairs of being able to offer an analysis of causal deviancy (Davidson 1973: 79). Many philosophers believe that the fate of the causalist account of action explanation depends on the availability of a purely causal account of this deviancy. But Davidson does not seem to suggest that his view is jeopardized by the impossibility of spelling out the conditions of deviancy in non-mentalistic and non-normative terms. After all, his view of the relation between psychological and causal laws seems to entail the impossibility of a purely causal account of deviancy, given that deviancy is a concern for psychology and rationality but not for mere causality.

What is the role of talk of intention in Davidson's account of intentional agency? According to Davidson's initial account, the term "intention" is *syncategorematic*; it has meaning only within a larger linguistic context. To say that the action  $\phi$  is performed *with the intention to K* only means that  $\phi$  is caused in the right way by a primary reason that includes a pro-attitude for actions of kind *K* (Davidson 1963: 8). The justification and explanation of intentional action only requires beliefs and pro-attitudes. On this view, performing  $\phi$  intentionally does not require that the agent have an intention to  $\phi$  as a *distinct* and *irreducible* mental state.

Davidson's early account of intentional agency has three main merits: (a) it combines rational justification with causal explanation; (b) it is ontologically parsimonious, since it does not countenance the existence of intentions as distinct mental states; and (c) it avoids the problems of traditional volitionism, since the intentionality of action does not depend on the contribution of an additional volitional attitude or mysterious acts of the will.

The account has two main limitations. First, it is silent on the deliberative process that might precede the acquisition of the pro-attitude and might handle conflicts among reasons for action. Second, it does not address the future-directed or prospective dimension of intentions, that is, the role that intentions might play in guiding temporally extended activities, controlling conduct in the distal future, framing future deliberation, and coordinating deliberation and action over time.

The prospective dimension of intentions is only accidental to the role Davidson attributes to the primary reason. The belief/pro-attitude pair might justify and explain an action in the distal future if the combination of the two attitudes initiates a causal chain that leads to a delayed execution of the action. Even so, the temporal distance between the acquisition of the primary reason and the action is not bridged by any kind of *guidance* by the primary reason.

A further prospective element might be located in the pro-attitude, since the agent might look favorably on the action on account of some of its distal effects. For instance, when the queen pours the poison in the king's ear in order to kill him, she does so out of pro-attitude in favor of the distal causal effects of her action. But the action that is rationalized and explained by this

future-directed pro-attitude is a *present* action – the pouring of the poison – although a present action that is intentional in light of some of its future causal effects. These effects are not necessarily produced by the continuous guidance of the primary reason. The queen’s agential control is over once the poison has been poured. At that point, “she has done her work; it only remains for the poison to do its” (Davidson 1971: 58).

Davidson’s initial account of intention pays limited attention to the upstream and downstream dimensions of intentional agency. Its main focus is on what might be called the “executive transition,” the transition from the presence of adequate reasons for action to the action’s inception. This transition amounts to the minimal and nonprospective form of the *settling* of a practical matter.

With respect to the executive transition, Davidson espouses a form of *volitional internalism*: once the primary reason is in place, no additional volitional attitude is necessary to give rise to the action (even if the action is at some causal distance from the primary reason). As a result, Davidson faces the standard problem of volitional internalism: how is clear-eyed weakness of will possible? How could one act intentionally contrary to one’s better judgment? Addressing this problem will induce some significant changes in Davidson’s account, as we are going to see in the discussion of Davidson (1969) later on.

## 2. Agency

Davidson (1971) discusses intentional action in relation to the metaphysical question of the distinction between actions and mere happenings. In this paper, Davidson expands on the account of action individuation already presented in Davidson (1963) while leaving untouched the account of intentional agency. (The paper does not hint at the revisions of the original account of intention contained in Davidson 1969, a paper that was published earlier but on which Davidson was working roughly around the same time as Davidson 1971 – see Davidson (1991: ix).)

Davidson first suggests accounting for the nature of agency by appealing to causality rather than to intentionality.

The question of the nature of agency is “greatly simplified” by formulating it in terms of *primitive actions* (Davidson 1971: 61). In terms of primitive actions, the problem of agency becomes the problem of accounting for the relation between the agent and the event that makes that event a primitive action.

Davidson first rejects the suggestion of appealing to *agent* causality. If the agent’s causing the action in the mode of agent causality is a distinct event, one faces the problem of traditional volitionism about “acts of will”: either they are actions, in which case we still need an account of their agency, or they are not, in which case we would be trying “to explain agency by appeal to an even more obscure notion, that of a causing that is not a doing” (Davidson 1971: 52).

Davidson turns next to *event* causality. This seems to be more promising given that there is a distinctive relationship that an action bears to its causal effects. Unlike all other events, actions can be described in terms of their causal effects, even if these effects are not part of the action itself. Only actions are subjected to the “accordion effect” (Davidson 1971: 53). This is what makes possible describing a finger movement as “turning on the light” or “alerting the prowler,” say. This feature of agency does not involve intention. *Any* causal effect of the action can be a legitimate description of it – this is why the alerting of the prowler, although unintentional, is as

acceptable as a description of the action as any of the other effects: “Once he has done one thing (move a finger), each consequence presents us with a deed; an agent causes what his actions cause” (Davidson 1971: 53). The agent’s deeds are not necessarily *intentional* ones.

The accordion effect, however, only offers a linguistic *mark* of agency, not an analysis of it (Davidson 1971: 52, 60). It speaks to the way in which agency (and possibly responsibility for it) can spread from primitive actions to their causal effects, but it says nothing about what makes a primitive action an action in the first place.

Davidson’s conclusion is that it is impossible to offer an analysis of agency without appealing to the concept of intention. As Davidson (1963) already stated, a person is the agent of an event only insofar as there is a description of the event (possibly in terms of some of its causal effects) under which it is true to say that the person did it *intentionally* (Davidson 1971: 46). The relation between the agent and the event is extensional (it is independent of how it is described), but its analysis relies on the concept of intention and its intensional character (Davidson 1971: 61).

For Davidson, this conclusion is somewhat “disappointing,” given his initial hope for an account of agency that would not appeal to the “more obscure” concept of intention (Davidson 1971: 47). In the concluding section of the paper, Davidson leaves open the question whether the concept of intention can be analyzed in simpler terms. Nevertheless, he hints at the availability of such an analysis in his earlier work. As seen earlier, Davidson (1963: 8) suggests that the term “intention” has meaning only as part of the locution “with the intention of”; a locution that is basic and refers to certain features of the action’s primary reason (viz., to say that an action  $\phi$  is performed with the intention to  $K$  means that  $\phi$  is caused in the right way by a primary reason that includes a pro-attitude for actions of kind  $K$ ). Although Davidson’s (1963) suggestion was not originally offered as an analysis of intentional action, it provides the basis for an account of the notion of intention in terms of the idea of “primary reason” and its role in causing actions. Hence, by combining the claims of Davidson (1963) and Davidson (1971), Davidson is able to provide an account of agency (both in its intentional and nonintentional aspects) in terms of efficient causality and without the need of intentions as distinct mental states.

Although “intentional” is only a qualification that applies to an action in virtue of the reason why one performs it, for Davidson, *all* actions are intentional (that is, under some description). For any given action, there is always a description that applies to it in the mode that makes that action intentional. Davidson does not discuss agency of a lesser and nonintentional kind, what other philosophers call “purposive” or “goal-directed” agency (such as the agency that might be attributed to nonhuman animals). Davidson only contrasts intentional agency to either events that are not actions *at all* or to the nonintentional *aspects* of intentional actions. His discussion is internal to the kind of agency that is distinctive (but not exhaustive) of the conduct of adult human beings. However, his account in terms of primary reasons might be a better fit for merely purposive action, which seems to require nothing more than motivation by the belief/pro-attitude pair in the way indicated by Davidson (Velleman 1992). This might explain some of the limitations of his initial account of intentionality, including the lack of attention to more complex forms of deliberation and to the prospective character of intentions.

### 3. Weakness of Will

Davidson (1969) amends the initial account of intentional action to make room for the possibility of akrasia or weakness of will – the possibility of acting intentionally against one’s better judgment. The belief/pro-attitude pair is no longer said to give rise to a primary reason that

causes the intentional action  $\phi$ . Rather, the combination of the two attitudes gives rise to a *conditional evaluative judgment* about the desirability of the particular action  $\phi$ . The agent judges that the action  $\phi$  is *prima facie* desirable. This means that the agent looks favorably on the action in light of some considerations but that, at least for the time being, she is not supporting it *unconditionally* or *no matter what*. The action, however, is supposed to follow only from an unconditional evaluative judgment: “Reasoning that stops at conditional judgments ... is practical only in its subject, not in its issue” Davidson (1969: 39).

A *prima facie* judgment in support of  $\phi$  is consistent with other contemporaneous *prima facie* evaluative judgments in support of actions that are incompatible with  $\phi$ -ing. But the agent cannot pursue incompatible actions; hence, she needs to reach an *unconditional* judgment in support of only one of the incompatible actions.

For Davidson, the rational transition to the unconditional judgment takes two distinct steps. First, the transition from the various conditional judgments to a single judgment, which Davidson calls an “*all things considered*” judgment (*ATC*). Second, a transition from the *ATC* judgment to an unconditional judgment that is supposed to lead to the action.

Consider the first transition. Starting from the *prima facie* judgments, the agent needs to form an “*all things considered*” (*ATC*) judgment, which can support only one of the incompatible actions. For instance, if  $\phi$  and  $\psi$  are incompatible actions and the agent makes the following two *prima facie* judgments:

<list>

PF1: action  $\phi$  of kind  $K$  is *prima facie* desirable.

PF2: action  $\psi$  of kind  $L$  is *prima facie* desirable.

</list>

On the basis of these two judgments, she needs to reach an *ATC* judgment that supports only one the two incompatible actions. Let us imagine that, *all things considered*, she judges that  $\phi$  is the desirable action (where the “things” to be taken into account are only the considerations that are accessible to the agent and deemed relevant by her (Davidson 1969: 40)).

According to Davidson (1969: 36), the formation of the *ATC* judgment is a matter of “weighing” the relevant considerations. His account of weakness of will, however, does not depend on a specific view about the sort of practical reasoning required in order to take into account all relevant considerations and determine which action to perform. What is crucial to his argument for the possibility of weakness of will is only that there is some form of practical reasoning that concludes into a single *ATC* judgment on the basis of the various *prima facie* judgments. (Notice that according to the current philosophical usage the conditional judgments should be better described as being *pro tanto* rather than *prima facie*. The expression “*prima facie*” might suggest that the considerations in support of the action might lose their justificatory force as the agent proceeds to combine the qualified judgments into the *ATC* one. But in ordinary practical reasoning, the *ATC* judgment does not normally *cancel* the force of *prima facie* considerations; it only indicates which among the competing pro-attitudes is going to be satisfied under the current circumstances.)

The *ATC* judgment is still conditional, even if conditional on all the things considered by the agent. As such, the *ATC* judgment does not yet issue into action. Action is supposed to follow



only from an “all-out,” unconditional judgment in support of the action (Davidson 1978: 98).

The all-out judgment takes up the role in the justification and explanation of intentional action that was earlier attributed to the primary reason. The belief/pro-attitude pair still contributes to justification and explanation, but only indirectly. First, the pair is an input to practical reasoning. Second, if it eventuates in an all-out judgment, the pair determines the descriptions under which the action is intentional (the ATC and all-out judgments do not provide their own descriptions, but rather inherits them from the prima facie judgment).

The rational transition from an ATC judgment to an all-out one is a matter of conforming to the “principle of continence,” which demands that one “perform the action judged best on the basis of all available relevant reasons” (Davidson 1969: 41). In other words, this principle requires that one form the unqualified judgment (and act out of it) simply by dropping the qualification from the ATC judgment.

The step from the ATC judgment to the all-out one makes weakness of will possible. If the agent fails to comply with the principle of continence, she derives the all-out judgment from one of the prima facie judgments rather than from the ATC one. The akratic agent judges all things considered that  $\phi$  is desirable but derives the all-out judgment that  $\psi$  is desirable directly from the prima facie judgment that  $\psi$  is desirable.

The performance of  $\psi$  counts as a genuine case of weakness of will because it is an intentional action (it is performed on the basis of an all-out judgment) contrary to one’s better judgment (the ATC judgment in support of  $\phi$ ).

Davidson’s account has been very influential in rekindling the philosophical discussion on the nature of weakness of will and in framing it in nonmoralized terms. The account has also been influential in presenting weakness of will as a problem of practical reasoning rather than the resultant of a mere conflict of desires. Nevertheless, Davidson has been far less successful in convincing other philosophers of the correctness of his positive account (for an introduction to the contemporary development of the discussion of weakness of will, see Stroud 2008).

Even if one were to accept the positive account, notice that for Davidson, weakness of will arises only in the transition from the ATC judgment to the all-out judgment. Could not there be a form of weakness of will that might strike *after* the all-out judgment has been acquired? This is possible to the extent that one claims that the all-out judgment is distinct both from the intention to act and the action itself. If so, it might be possible to imagine a form of weakness of will that consists in the acquisition of akratic intentions (intentions that go against one’s all-out judgments, rather than against one’s ATC judgments), or in the failure to acquire the intention to  $\phi$  in spite of one’s all-out judgment in favor of  $\phi$ -ing. Or one might imagine cases in which one fails to act in accordance with one’s all-out judgment. This might seem especially plausible when there is a time interval between the all-out judgment in favor of  $\phi$ -ing and the time when one is to  $\phi$ .

Davidson does not explicitly discuss these scenarios, but his view has the resources to deny at least the possibility of akratic intentions and the failure to acquire intentions in accordance with one’s all-out judgments. In order to do so, Davidson would have to identify the intention with the all-out judgment – which is the position that Davidson comes to embrace in his later work on intention; see Davidson (1978: 99).

What is the import of the account of weakness of will on Davidson’s view of intentional agency as presented in Davidson (1969)?

First, acting intentionally is no longer equated with being caused by a primary reason as a belief–pro-attitude pair. Rather, one acts intentionally when the action is caused by the all-out judgment of the desirability of that action. Up to this point, he does not need to countenance intentions as distinct states of mind. But there is something distinctive about this judgment of the desirability of the action: its *all-out* character. Davidson has come to acknowledge that the executive transition requires a kind of settling of the practical matter that is not secured by the mere combination of a belief and a pro-attitude. Settling a practical matter now calls for two steps: (a) the coming to a close of the deliberation – that is, the acquisition of the ATC judgment; (b) the move toward action – that is, the dropping of the ATC qualification and the acquisition of the all-out judgment, which makes the judgment “practical in its issue” (Davidson 1969: 39).

Second, the account is still compatible with volitional internalism. No additional volitional attitude is introduced to account for the executive transition, for the move from the acquisition of all-out judgment about the action’s desirability to the action itself. One might wonder whether the compliance with the principle of continence (which presides over the transition from the ATC to the all-out judgment) might not require a novel attitude, such as the endorsement of the principle, which the weak-willed agent might either lack or fail to be moved by. According to Davidson’s view of the principles of rationality, however, there is no need for such additional attitude. For him, the principles of rationality describe *patterns* of attitudes too much deviation from which undermines the possibility of continuing to describe their subject as a rational agent (see Chapter 10).

Third, the pro-attitudes involved in intentional agency are now explicitly presented as *evaluative judgments*. In the earlier account, they were characterized more broadly, since they encompassed conative attitudes – such as urges and promptings – that might have no evaluative dimension (although they indicate “what it was about the action that appealed” to the agent (Davidson 1963: 3–4)). In their guise as evaluative judgments, the pro-attitudes are a better fit to work as premises and conclusions of practical deliberation.

Although he does not explicitly discuss the relation between intentions and all-out judgments, the special role of these judgments might make one wonder whether they might be closely related to intentions. Davidson does not raise this question in the paper, but he eventually comes to argue for the *identification* of intentions with all-out evaluative judgments, as we are going to see now.

## 4. Intending

In his last sustained discussion of intentional agency in Davidson (1978), Davidson changes his mind about the nature and existence of intentions. He comes to the conclusion that intentions are psychological attitudes of a *distinct* and *irreducible* kind.

Davidson begins his discussion by introducing *pure intending*: “an intending that may occur without practical reasoning, action, or consequence” (Davidson 1978: 83). He offers the following illustration:

Someone may intend to build a squirrel house without having decided to do it, deliberated about it, formed an intention to do it, or reasoned about it. And despite his intention, he may never build a squirrel house, try to build a squirrel house, or do anything whatever with the intention of getting a squirrel house built. (Davidson 1978: 83)

Pure intendings cannot be accommodated by the earlier account of intentional agency in Davidson (1963) and Davidson (1971). But one might wonder whether their isolation from both

deliberation and action make them irrelevant for the understanding of the nature of intentional action. In response, Davidson claims that although one might imagine that intentional action could occur without the prior extra element of pure intending, “it would be *astonishing* if that extra element were foreign to our understanding of intentional action” (Davidson 1978: 88, my emphasis; see also 89). Consider a temporally extended action  $\psi$ . As the agent takes any of the momentary steps necessary to carry it out at any particular moment, one continues to have an intention toward the complete action that is still in progress. It would be surprising, Davidson argues, if this future-directed intention, which the agent has throughout the execution of the extended action  $\psi$ , were not the same as the pure intention to  $\psi$  that he had prior to the inception of the action (Davidson 1978: 88, 96).

Davidson goes on to argue that this distinct component of intentional action is *irreducible* to any of the elements that are already available in his earlier account (actions, beliefs, and desires).

Can intention be an action? Pure intending cannot be an action because it is a mental *state* rather than an event. The acquisition of the intention, on the other hand, is an event and thus might be an action. But this claim offers “little illumination” about the nature of intention (Davidson 1978: 89).

A performative theory of intentions might seem more promising. It would equate forming an intention to the issuing of a self-directed promise or command. But this account does not work because, unlike intentions, promises and commands involve highly specific conventions and public performances. In addition, they generate obligations that persist even when one changes one’s mind, whereas “if an agent does not do what he intended to do, he does not normally owe himself an explanation or apology, especially if he simply changed his mind” (Davidson 1978: 90).

Davidson goes on to argue that the intention to  $\phi$  cannot be identified with the belief that one will  $\phi$ . To begin with, he presents a case where one can do something intentionally without believing that one is doing it:

In writing heavily on this page I may be intentioned to produce ten legible carbon copies. I do not know, or believe with any confidence, that I am succeeding. But if I am producing ten legible carbon copies, I am certainly doing it intentionally. (Davidson 1978: 92)

A similar distinction holds of future-directed intentions: I can intend to do something in the future without believing that I will do it.

One might reply that the distinction is only apparent because the intention is not accurately stated. Given the uncertainty about one’s future conduct, one only believes that one is going to *try* to  $\phi$ . The intention can be identical with this belief if what one really intends is to *try* to  $\phi$  rather than to  $\phi$ . Davidson’s response is that although in some cases it might be appropriate to claim that one just intends to try, this is not true of intentions in general.

The reduction of intention to belief might be supported by claiming that ordinary statements of intention are usually *elliptical*: an intention to  $\phi$  usually omits to specify the conditions under which one does not expect to  $\phi$ . If those conditions are made explicit, the intention has the same content as one’s belief about one’s future conduct, which makes the identification of the intention with that belief possible.

Davidson sees an insurmountable problem with this suggestion. It is impossible to specify *all* the conditions that stand in the way of one’s future performance since “there can be no finite list of things we think might prevent us from doing what we intend, or circumstances that might cause

us to stay our hand” (Davidson 1978: 94). To get around this, one might try to use a general and generic qualification, such as “I intend to  $\phi$  if nothing prevents me, if I don’t change my mind, and if nothing untoward happens.” But this would not help with the identification of the intention with the belief, since the corresponding belief – the belief that I will  $\phi$  if nothing prevents me if I don’t change my mind, and if nothing untoward happens – would “tell us almost nothing about what the agent believes about the future” (Davidson 1978: 93). A generic qualification of this kind would be in terms of what Davidson (1978: 94) deems to be “bogus conditions.”

Davidson does not deny that there could be genuinely conditional intentions. These intentions only qualify the action in terms of a *limited* set of *specific* circumstances: the conditions that the agent *explicitly* considers in acquiring the intention and that count as “bona fide.” They are conditions that refer to “reasons for acting that are *contemporaneous* with the intention.” (Davidson 1978: 94, my emphasis) These are conditions on the *advisability* of the intention and the action, rather than on their *possibility*. The latter conditions are not reasons for acting that could be contemporaneous with the intention to  $\phi$ . They are *enabling* conditions for intending and acting. If the agent knows that they are not going to hold, she cannot intend to  $\phi$ , since the intention is either impossible or irrational on account of the impossibility of her doing  $\phi$ .

The bona fide conditions, however, fall far short of specifying *all* the circumstances under which one expects not to  $\phi$  even if one intends to do it. Hence, the intention to  $\phi$  cannot be identical to a belief that one will  $\phi$ , even if one were to state the intention in more accurate conditional terms (Davidson 1978: 94–95). (For additional discussion of Davidson’s account of conditional intentions, see Ferrero 2009.)

The “strongest argument” against the reduction of intention to belief is the difference in the reasons that support the two kinds of attitude. Whereas the intention expressed by “I will do  $\phi$ ” is supported by the kinds of considerations advanced in first-person practical reasoning, the belief that I will do  $\phi$  requires “additional assumptions,” including beliefs about one’s psychological states and the working of one’s psychology, from which one can derive third-personal predictions about one’s future conduct (Davidson 1978: 95).

The “most promising” attempt at the reduction of intentions is in terms of desires/wants, given that both intentions and wants are kinds of pro-attitudes. But desires are only *prima facie* judgments and, as such, “cannot be directly associated to actions, for its not reasonable to perform an action merely because it has a desirable characteristic” (Davidson 1978: 98). The performance of an action requires a *further* judgment that “the desirable characteristic was enough to act on” (Davidson 1978: 98). This further judgment is the all-out one. If the execution of the desirable action is immediately possible, the judgment might be identical with the action itself. When the action lies in the future, the judgment is the same as the future-directed intention.

But how can the judgment be about a particular action, given that the action lies in the future and the agent is not yet acquainted with it? The all-out future-directed judgment is actually about a *set* of actions, but it is more specific than the *prima facie* judgment on the desirability of actions of kind *K*. It would be “mad” to approve of and set oneself to perform *any* future action just because it is of kind *K*, regardless of the circumstances (Davidson 1978: 99). But given one’s current beliefs about the future, which exclude circumstances that would make an action of kind *K* impossible or inadvisable, one does approve of any of the actions of kind *K* that one expects to be able to perform at the intended future time.

The role of this “view of the future” explains two features of intentions. First, it explains why the intention is not conditional on the circumstances that make it possible and reasonable. The

intention is rather *conditioned* by the beliefs about its possibility and advisability. These beliefs are conditions on the adoption of the intention, rather than qualifiers of its content. Second, it explains why the intention to  $\phi$  usually goes together with (but does not reduce to) the belief that one will  $\phi$  (Davidson 1978: 100).

In his later account of intentions, Davidson offers a sustained and plausible case for the existence of intentions as irreducible mental states. Most of his argument aims at rejecting attempts at reduction. Davidson ends up saying little about the positive nature of intentions. His discussion of their role in practical reasoning complements the earlier discussion of all-out judgments in Davidson (1969). He offers some novel considerations about the role of intentions as directed to *future* actions. In particular, he claims that an intention for an extended action continues to persist as the same attitude from the inception of the action throughout its unfolding, and moves the agent at each given time to take whatever step is required at that time. The same intention also exists *prior* to the inception of the action as a pure intending. And the inception of the action is explained by the agent coming to believe that the time of the intended action has now arrived (Davidson 1978: 88, 96).

This account still leaves some open questions. To begin with, what does explain the possibility of a pure intention that *never* issues into action? If I intend to build a squirrel house but I never try to build it or take any preparatory steps, is it just because I never come to believe that it is now the time to act? This explanation would fit Davidson's account if my intention to build the squirrel house does not specify the time of action ("I intend to build a squirrel house *one day*"). But is it possible to have a pure intention to do something at a particular time, come to believe that the time is now, and yet not be moved to action? Failures to carry out an occurrent intention to  $\phi$  when one believes that now is the time to  $\phi$  are not unheard of. Some of these failures seem to be instances of weakness of will, but they cannot be accommodated by Davidson's (1969) account. According to that account, weakness of will is a matter of the violation of the principle of continence in the transition from an ATC judgment to an all-out judgment. But in this case, the problem arises with the transition from the all-out judgment (the pure intending) to the action.

The identification of intentions with all-out judgments speaks to what might be called the upstream dimension of intentions, that is, to their justification and to their relation to the practical reasoning that might lead to their acquisition (notice that an actual deliberation is not necessary to acquire an intention; see Davidson 1978: 85). But by identifying the intentions with all-out evaluative judgments, Davidson is unable to account for intentions that we seemingly acquire independently of, if not even contrary to, our evaluative judgments, including intentions that might be spontaneously acquired for no reason, and those acquired by choosing a particular action over an equally choiceworthy alternative one in cases of normative underdetermination (see Bratman 1999: 219).

Finally, what about the downstream dimension of intentions – their role in guiding temporally extended activities, controlling future action, coordinating conduct and deliberation over time? Davidson only focuses on the guiding role of a future-directed intention in initiating action and in carrying it out over time. He discusses how incompatible actions are ruled out at the deliberative stage (when one is to reach a single ATC judgment in the face of *prima facie* judgments in support of incompatible actions), but he does not explicitly address the role that an intention, once adopted, an intention plays in framing *further* deliberation and in coordinating with other actions over time (for this criticism, see Bratman 1999).

## References

- <book>Bratman, M. (1999). Davidson's theory of intentions. In M. Bratman, *Faces of Intention*, Chapter 11. Cambridge, UK: Cambridge University Press, pp. 209–224.</book>
- <book>Davidson, D. (1963). Action, reasons, and causes. In D. Davidson, *Essays on Actions and Events*, Chapter 1. Oxford: Oxford University Press, pp. 3–20.</book>
- <book>———. (1969). How is weakness of the will possible? In D. Davidson, *Essays on Actions and Events*. Oxford: Oxford University Press, pp. 21–42.</book>
- <book>———. (1971). Agency. In D. Davidson, *Essays on Actions and Events*, Chapter 3,. Oxford: Oxford University Press, pp. 43–61.</book>
- <book>———. (1973). Freedom to act. In D. Davidson, *Essays on Actions and Events*, Chapter 4. Oxford: Oxford University Press, pp. 63–82.</book>
- <book>———. (1978). Intending. In D. Davidson, *Essays on Actions and Events*, Chapter 5,. Oxford: Oxford University Press, pp. 83–102.</book>
- <book>———. (1991). *Essays on Actions and Events*. Oxford: Oxford University Press.</book>
- <journal>Ferrero, L. (2009). Conditional intentions. *Noûs* 43(4):700–741.</journal>
- <book>Lepore, E. and Ludwig, K. (eds). (2013). *Companion to Davidson*. Oxford: Wiley-Blackwell.</book>
- <book>Stroud, S. (2008). Weakness of will. In *Stanford Encyclopedia of Philosophy*, E.N. Zalta (ed.). Stanford, CA: Stanford University. <http://plato.stanford.edu/entries/weakness-will/> (accessed •February 8, 2013).</book>
- <journal>Velleman, J.D. (1992). The guise of the good. *Noûs* 26(1):3–26.</journal>