If a person's head moves, she may or may not have moved her head, and, if she did move it, she may have actively performed the movement of her head or merely, by doing something else, caused a passive movement. And, if she performed the movement, she might have done so intentionally or not. This short array of contrasts (and others like them) has motivated questions about the nature, variety, and identity of action. Beyond the matter of her moving, when the person moves her head, she may be indicating agreement or shaking an insect off her ear. Should we think of the consequences, conventional or causal, of physical behavior as constituents of an action distinct from but ‘generated by’ the movement? Or should we think that there is a single action describable in a host of ways? Also, actions, in even the most minimal sense, seem to be essentially ‘active’. But how can we explain what this property amounts to and defend our wavering intuitions about which events fall in the category of the ‘active’ and which do not?

Donald Davidson [1980, essay 3] asserted that an action, in some basic sense, is something an agent does that was ‘intentional under some description,’ and many other philosophers have agreed with him that there is a conceptual tie between genuine action, on the one hand, and intention, on the other. However, it is tricky to explicate the purported tie between the two concepts. First, the concept of ‘intention’ has various conceptual inflections whose connections to one another are not at all easy to delineate, and there have been many attempts to map the relations between intentions for the future, acting intentionally, and acting with a certain intention. Second, the notion that human behavior is often intentional under one description but not under another is itself hard to pin down. For example, as Davidson pointed out, an agent may intentionally cause himself to trip, and the activity that caused the tripping may have been intentional under that description while, presumably, the foreseen but involuntary tripping behavior that it caused is not supposed to be intentional under any heading. Nevertheless, both the tripping and its active cause are required to make it true that the agent intentionally caused himself to trip. Both occurrences fall equally, in that sense, ‘under’ the operative description. So further clarification is called for.

There has been a notable or notorious debate about whether the agent's reasons in acting are causes of the action — a longstanding debate about the character of our common sense explanations of actions. Some philosophers have maintained that we explain why an agent acted as he did when we explicate how the agent's normative reasons rendered the action intelligible in his eyes. Others have stressed that the concept of ‘an intention with which a person acted’ has a teleological dimension that does not, in their view, reduce to the concept of ‘causal guidance by the agent's reasons.’ But the view that reason explanations are somehow causal explanations remains the dominant position. Finally, recent discussions have revived interest in important questions about the nature of intention and its distinctiveness as a mental state, and about the norms governing rational intending.

1. The Nature of Action and Agency
   1.1 Knowledge of one's own actions.
   1.2 Governance of one's own actions.
2. Intentional Action and Intention
3. The Explanation of Action
4. Intentions and Rationality
Bibliography
Academic Tools
Other Internet Resources
Related Entries
It has been common to motivate a central question about the nature of action by invoking an intuitive distinction between the things that merely happen to people — the events they undergo — and the various things they genuinely do. The latter events, the doings, are the acts or actions of the agent, and the problem about the nature of action is supposed to be: what distinguishes an action from a mere happening or occurrence? For some time now, however, there has been a better appreciation of the vagaries of the verb ‘to do’ and a livelier sense that the question is not well framed. For instance, a person may cough, sneeze, blink, blush, and thrash about in a seizure, and these are all things the person has, in some minimal sense, ‘done,’ although in the usual cases, the agent will have been altogether passive throughout these ‘doings.’ It is natural to protest that this is not the sense of ‘do’ the canny philosopher of action originally had in mind, but it is also not so easy to say just what sense that is. Moreover, as Harry Frankfurt [1978] has pointed out, the purposeful behavior of animals constitutes a low-level type of ‘active’ doing. When a spider walks across the table, the spider directly controls the movements of his legs, and they are directed at taking him from one location to another. Those very movements have an aim or purpose for the spider, and hence they are subject to a kind of teleological explanation. Similarly, the idle, unnoticed movements of my fingers may have the goal of releasing the candy wrapper from my grasp. All this behavioral activity is ‘action’ in some fairly weak sense.

Nevertheless, a great deal of human action has a richer psychological structure than this. An agent performs activity that is directed at a goal, and commonly it is a goal the agent has adopted on the basis of an overall practical assessment of his options and opportunities. Moreover, it is immediately available to the agent's awareness both that he is performing the activity in question and that the activity is aimed by him at such-and-such a chosen end. At a still more sophisticated conceptual level, Frankfurt [1988, 1999] has also argued that basic issues concerning freedom of action presuppose and give weight to a concept of ‘acting on a desire with which the agent identifies.’ Under Frankfurt's influence on this point, a good deal has been written to elucidate the nature of ‘full-blooded’ human agency, whether the action is finally delineated either in Frankfurt's way or along different but related lines [see Velleman 2000, essay 6, Bratman 1999, essay 10]. Thus, there are different levels of action to be distinguished, and these include at least the following: unconscious and/or involuntary behavior, purposeful or goal directed activity (of Frankfurt's spider, for instance), intentional action, and the autonomous acts or actions of self-consciously active human agents. Each of the key concepts in these characterizations raises some hard puzzles.

### 1.1 Knowledge of one's own actions.

It is frequently noted that the agent has some sort of immediate awareness of his physical activity and of the goals that the activity is aimed at realizing. In this connection, Elizabeth Anscombe [1963] spoke of ‘knowledge without observation.’ The agent knows ‘without observation’ that he is performing certain bodily movements (perhaps under some rough but non-negligible description), and he knows ‘without observation’ what purpose(s) the behavior is meant to serve [see also Falvey 2000]. Anscombe's discussion of her claim is rich and suggestive, but her conception of ‘knowledge through observation’ is problematic. Surely, one wants to say, proprioception and kinesthetic sensation play some role in informing the agent of the positions and movements of his body, and it is uncertain why these informational roles should fail to count as modes of inner ‘observation’ of the agent's own overt physical behavior. What Anscombe explicitly denies is that agents generally know of the positions or movements of their own bodies by means of ‘separably describable sensations’ that serve as criteria for their judgements about the narrowly physical performance of their bodies. However, when a person sees that there is a goldfinch in front of him, his knowledge is not derived as an inference from the ‘separably describable' visual impressions he has in seeing the goldfinch, but this is an instance of knowledge through observation nonetheless.

In a related vein, David Velleman [1989] describes knowledge of one's present and incipient actions as ‘spontaneous’ (knowledge that the agent has achieved without deriving it from evidence adequate to warrant it), and as ‘self-fulfilling’ (expectations of acting that tend to produce actions of the kind expected). For Velleman, these expectations are themselves intentions, and they are chiefly derived by the agent through practical reasoning about what she is to perform. Thus, Velleman is what Sarah Paul (2009) calls a Strong Cognitivist, i.e., someone who identifies an intention with a certain pertinent belief about what she is doing or about to do. Setiya (2009) holds a similar view. A Weak Cognitivist, in Paul's terminology, is a theorist that holds that intentions to F are partially constituted by but are not identical with relevant beliefs that one will F. For instance, Paul Grice (1971) held that an intention to F consisted in the agent’s ‘willing’ herself to F combined with a belief that she will actually F as a more or less immediate consequence of her having so willed. Because Strong Cognitivists maintain that the intention/beliefs of the agent are predominantly not based either on observation or evidence of any sort, and because they claim in addition that these states are causally reliable in producing actions that validate their contents, such theorists believe that these intentions, when they have been carried out, constitute a mode of ‘practical’ knowledge that has not been derived from observation. Weak Cognitivists can construct a similar story about how the agent's own actions can, in a plausible sense, be known without relying on observation.

However, it is not obvious that an agent's knowledge of her intentional actions is not inferred from immediate knowledge of her own intentions. Consider, to illustrate the line of thought, Grice's theory of intention and belief. As noted above, he held a Weak Cognitivist view according to which an agent wills that he Fs and derives from his awareness of willing that he will in fact F (or at least try to F) precisely because he has willed to do so. However, it seems plausible, as Sarah Paul argues at length in her 2009, that intentions to F, rightly understood, can take the place of the counterpart ‘willings’ in Grice's account. Thus, an agent, intending to F in the near future, and being immediately aware of so intending, forms inferentially the belief that she will F soon (or at least try to F) precisely because she has intended to do so. After all, the conditional,
If the agent intends to \( F \) shortly and does not change her mind, then shortly she will at least try to \( F \).

appears to be knowable \( a \) priori. The belief that the agent thereby derives is, although it is inferred, not derived from observation. Paul labels this the “the inferentialist account,” and it is not easily ruled out. [See also Wilson 2000 and Moran 2001.] These puzzles about the nature of an agent's knowledge of her own intentional actions are thus closely intertwined with questions about the nature of intention and about the nature of the explanation of action. In the final section, we address briefly some further key issues that arise in this connection.

### 1.2 Governance of one's own actions.

It is also important to the concept of ‘goal directed action’ that agents normally implement a kind of direct control or guidance over their own behavior. An agent may guide her paralyzed left arm along a certain path by using her active right arm to shove it through the relevant trajectory. The moving of her right arm, activated as it is by the normal exercise of her system of motor control, is a genuine action, but the movement of her left arm is not. That movement is merely the causal upshot of her guiding action, just as the onset of illumination in the light bulb is the mere effect of her action when she turned on the light. The agent has direct control over the movement of the right arm, but not over the movement of the left. And yet it is hardly clear what ‘direct control of behavior’ can amount to here. It does not simply mean that behavior \( A \), constituting a successful or attempted \( F \), was initiated and causally guided throughout its course by a present-directed intention to be \( F \). Even the externally guided movement of the paralyzed left arm would seem to satisfy a condition of this weak sort. Alfred Mele [1992] has suggested that the intuitive ‘directness’ of the guidance of action \( A \) can partially be captured by stipulating that the action-guiding intention must trigger and sustain \( A \) proximally. In other words, it is stipulated that the agent's present-directed intention to be \( F \) should govern action \( A \), but not by producing some other prior or concurrent action \( A' \) that causally controls \( A \) in turn. But the proposal is dubious. On certain assumptions, most ordinary physical actions are liable to flunk this strengthened requirement.

The normal voluntary movements of an agent's limbs are caused by complicated contractions of suitable muscles, and the muscle contractions, since they are aimed at causing the agent's limbs to move, may themselves count as causally prior human actions. For instance, on Davidson's account of action they will since the agent's muscle contracting is intentional under the description ‘doing something that causes the arm to move’ [see Davidson 1980, essay 2]. Thus, the overt arm movement, in a normal act of voluntary arm moving, will have been causally guided by a prior action, the muscle contracting, and consequently the causal guidance of the arm's movement will fail to be an instance of ‘proximal’ causation at all [see Sehon 1998].

As one might imagine, this conclusion depends upon how an act of moving a part of one's body is to be conceived. Some philosophers maintain that the movements of an agent's body are never actions. It is only the agent's direct moving of, say, his leg that constitutes a physical action; the leg movement is merely caused by and/or incorporated as a part of the act of moving [see Hornsby 1980]. This thesis re-opens the possibility that the causal guidance of the moving of the agent's leg by the pertinent intention is proximal after all. The intention proximally governs the moving, if not the movement, where the act of moving is now thought to start at the earliest, inner stage of act initiation. Still, this proposal is also controversial. For instance, J.L. Austin [1962] held that the statement

(1) The agent moved his leg

is ambiguous between (roughly)

(1') The agent caused his leg to move

and the more specific

(1'') The agent performed a movement with his leg.

If Austin is right about this, then the nominalization “the agent's moving of his leg” should be correspondingly ambiguous, with a second reading that denotes a certain leg movement, a movement the agent has performed. Thus, no simple appeal to a putative distinction between ‘movement’ and ‘moving’ will easily patch up the conception of ‘direct control of action’ under present scrutiny.

In any event, there is another well-known reason for doubting that the ‘directness’ of an agent's governance of his own actions involves the condition of causal proximity — that an action is not to be controlled by still another action of the same agent. Some philosophers believe that the agent's moving his leg is triggered and sustained by the agent's trying to move his leg in just that way, and that the efficacious trying is itself an action [see Hornsby 1980, Ginet 1990, and O'Shaughnessy 1973, 1980]. If, in addition, the agent's act of leg moving is distinct from the trying, then, again, the moving of the leg has not been caused proximally by the intention. The truth or falsity of this third assumption is linked with a wider issue about the individuation of action that has also been the subject of elaborate discussion.

Donald Davidson [1980, essay 1], concurring with Anscombe, held that

(2) If a person \( Fs \) by \( G \), then her act of \( F = G \) = her act of \( G \).

In Davidson's famous example, someone alerts a burglar by illuminating a room, which he does by turning on a light, which he does in turn by flipping the appropriate switch. According to the Davidson/Anscombe thesis above, the alerting of the burglar = the illuminating of the room = the turning on of the light = the flipping of the switch. And this is so
despite the fact that the alerting of the burglar was unintentional while the flipping of the switch, the turning on of the light, and the illuminating of the room were intentional. Suppose now that it is also true that the agent moved his leg by trying to move his leg in just that matter. Combined with the Davidson/Anscombe thesis about act identification, this implies that the agent's act of moving his leg = his act of trying to move that leg. So, perhaps the act of trying to move the leg doesn't cause the act of moving after all, since they are just the same.

The questions involved in these debates are potentially quite confusing. First, it is important to distinguish between phrases like

(a) the agent's turning on the light

and gerundive phrases such as

(b) the agent's turning on of the light.

Very roughly, the expression (a) operates more like a 'that' clause, viz.

(a') that the agent turned on the light,

while the latter phrase appears to be a definite description, i.e.,

(b') the turning on of the light [performed] by the agent.

What is more, even when this distinction has been drawn, the denotations of the gerundive phrases often remain ambiguous, especially when the verbs whose nominalizations appear in these phrases are causatives. No one denies that there is an internally complex process that is initiated by the agent's switch-flipping hand movement and that is terminated by the light's coming on as a result. This process includes, but is not identical with, the act that initiates it and the event that is its culminating upshot. Nevertheless, in a suitable conversational setting, the phrases (b) and (b') can be properly used to designate any of the three events: the act that turned on the light, the onset of illumination in the light, and whole process whereby the light has come to be turned on. [For further discussion, see Parsons 1990, Pietrofsky 2000, and Higginbotham 2000].

Now, the Davidson-Anscombe thesis plainly is concerned with the relation between the agent’s act of turning on the light, his act of flipping the switch, etc. But which configuration of events, either prior to or contained within the extended causal process of turning on the light, really constitutes the agent's action? Some philosophers have favored the overt arm movement the agent performs, some favor the extended causal process he initiates, and some prefer the relevant event of trying that precedes and ‘generates’ the rest. It has proved difficult to argue for one choice over another without simply begging the question against competing positions. As noted before, Hornsby and other authors have pointed to the intuitive truth of

(3) The agent moved his arm by trying to move his arm,

and they appeal to the Davidson-Anscombe thesis to argue that the act of moving the arm = the act of trying to move the arm. On this view, the act of trying — which is the act of moving — causes a movement of the arm in much the same way that an act of moving the arm causes the onset of illumination in the light. Both the onset of illumination and the overt arm movement are simply causal consequences of the act itself, the act of trying to move his arm in just this way. Further, in light of the apparent immediacy and strong first person authority of agents' judgements that they have tried to do a certain thing, it appears that acts of trying are intrinsically mental acts. So, a distinctive type of mental act stands as the causal source of the bodily behavior that validates various physical re-descriptions of the act.

And yet none of this seems inevitable. It is arguable that

(4) The agent tried to turn on the light

simply means, as a first approximation at least, that

(4') The agent did something that was directed at turning on the light.

Moreover, when (4) or (4') is true, then the something the agent did that was directed at turning on the light will have been some other causally prior action, the act of flipping the switch, for example. If this is true of trying to perform basic acts (e.g., moving one's own arm) as well as non-basic, instrumental acts, then trying to move one's arm may be nothing more than doing something directed at making one's arm move. In this case, the something which was done may simply consist in the contracting of the agent's muscles. Or, perhaps, if we focus on the classic case of the person whose arm, unknown to her, is paralyzed, then the trying in that case (and perhaps in all) may be nothing more than the activation of certain neural systems in the brain. Of course, most agents are not aware that they are initiating appropriate neural activity, but they are aware of doing something that is meant to make their arms move. And, in point of fact, it may well be that the something of which they are aware as a causing of the arm movement just is the neural activity in the brain. From this perspective, 'trying to F' does not name a natural kind of mental act that ordinarily sets off a train of fitting physical responses. Rather, it gives us a way of describing actions in terms of a goal aimed at in the behavior without committing us as to whether the goal was realized or not. It also carries no commitment,
i. concerning the intrinsic character of the behavior that was aimed at \( F \),
ii. whether one or several acts were performed in the course of trying, and
iii. whether any further bodily effects of the trying were themselves additional physical actions [see Cleveland 1997].

By contrast, it is a familiar doctrine that what the agent does, in the first instance, in order to cause his arm to move is to form a distinctive mental occurrence whose intrinsic psychological nature and content is immediately available to introspection. The agent \( \text{wills} \) his arm to move or produces a \( \text{volition} \) that his arm is to move, and it is this mental willing or volition that is aimed at causing his arm to move. Just as an attempt to turn on the light may be constituted by the agent's flipping of the switch, so also, in standard cases, trying to move his arm is constituted by the agent's willing his arm to move. For traditional 'volitionalism,' willings, volitions, basic tryings are, in Brian O'Shaughnessy's apt formulation, 'primitive elements of animal consciousness.' They are elements of consciousness in which the agent has played an active role, and occurrences that normally have the power of producing the bodily movements they represent. Nevertheless, it is one thing to grant that, in trying to move one's body, there is some 'inner' activity that is meant to initiate an envisaged bodily movement. It is quite another matter to argue successfully that the initiating activity has the particular mentalistic attributes that volitionalism has characteristically ascribed to acts of willing.

It is also a further question whether there is only a single action, bodily or otherwise, that is performed along the causal route that begins with trying to move and terminates with a movement of the chosen type. One possibility, adverted to above, is that there is a whole causal chain of actions that is implicated in the performance of even the simplest physical act of moving a part of one's body. If, for example, 'action' is goal-directed behavior, then the initiating neural activity, the resulting muscle contractions, and the overt movement of the arm may all be actions on their own, with each member in the line-up causing every subsequent member, and with all of these actions causing an eventual switch flipping somewhere further down the causal chain. On this approach, there may be nothing which is the act of flipping the switch or of turning on the light, because each causal link is now an act which flipped the switch and (thereby) turned on the light [see Wilson 1989]. Nevertheless, there still will be a single overt action that made the switch flip, the light turn on, and the burglar become alert, i.e., the overt movement of the agent's hand and arm. In this sense, the proposal supports a modified version of the Davidson/Anscombe thesis.

However, all of this discussion suppresses a basic metaphysical mystery. In the preceding two paragraphs, it has been proposed that the neural activity, the muscle contractions, and the overt hand movements may all be actions, while the switch's flipping on, the light's coming on, and the burglar's becoming alert are simply happenings outside the agent, the mere effects of the agent's overt action. As we have seen, there is plenty of disagreement about where basic agency starts and stops, whether within the agent's body or somewhere on its surface. There is less disagreement that the effects of bodily movement beyond the body, e.g., the switch's flipping on, the onset of illumination in the room, and so on, are not, by themselves at least, purposeful actions. Still, what could conceivably rationalize any set of discriminations between action and non-action as one traces along the pertinent complex causal chains from the initial mind or brain activity, through the bodily behavior, to the occurrences produced in the agent's wider environment?

Perhaps, one wants to say, as suggested above, that the agent has a certain kind of direct (motor) control over the goal-seeking behavior of his own body. In virtue of that fundamental biological capacity, his bodily activity, both inner and overt, is governed by him and directed at relevant objectives. Inner physical activity causes and is aimed at causing the overt arm movements and, in turn, those movements cause and are aimed at causing the switch to flip, the light to go on, and the room to become illuminated. Emphasizing considerations of this sort, one might urge that they validate the restriction of action to events in or at the agent's body. And yet, the stubborn fact remains that the agent also does have a certain 'control' over what happens to the switch, the light, and even over the burglar's state of mind. It is a goal for the agent of the switch's flipping on that it turn on the light, a goal for the agent of the onset of illumination in the room that it render the room space visible, etc. Hence, the basis of any discrimination between minimal agency and non-active consequences within the extended causal chains will have to rest on some special feature of the person's guidance: the supposed 'directness' of the motor control, the immediacy or relative certainty of the agent's expectations about actions vs. results, or facts concerning the special status of the agent's living body. The earlier remarks in this section hint at the serious difficulty of seeing how any such routes are likely to provide a rationale for grounding the requisite metaphysical distinction(s).

2. Intentional Action and Intention

Anscombe opened her monograph *Intention* by noting that the concept of 'intention' figures in each of the constructions:

(5) The agent intends to \( G \);

(6) The agent \( G \)'d intentionally; and

(7) The agent \( F \)'d with the intention of \( G \).

For that matter, one could add

(7') In \( F \)'ing (by \( F \)), the agent intended to \( G \).

Although (7) and (7') are closely related, they seem not to say quite the same thing. For example, although it may be true that
Despite the differences between them, we will call instances of (7) and (7*) ascriptions of intention in action. These sentential forms represent familiar, succinct ways of explaining action. A specification of the intention with which an agent acted or the intention that the agent had in acting provides a common type of explanation of why the agent acted as he did. This observation will be examined at some length in Section 3.

Statements of form (5) are ascriptions of intention for the future, although, as a special case, they include ascriptions of present-directed intentions, i.e., the agent's intention to be G now. Statements of form (6), ascriptions of acting intentionally, bear close connections to corresponding instances of (7). As a first approximation at least, it is plausible that (6) is true just in case

(6') The agent G'd with the intention of (thereby) G'ing.

However, several authors have questioned whether such a simple equivalence captures the special complexities of what it is to G intentionally. Here is an example adapted from Davidson [1980, essay 4]. Suppose that Betty kills Jughead, and she does so with the intention of killing him. And yet suppose also that her intention is realized only by a wholly unexpected accident. The bullet she fires misses Jughead by a mile, but it dislodges a tree branch above his head and releases a swarm of hornets that attack him and sting him until he dies. In this case, it is at least dubious that, in this manner, Betty has killed Jughead intentionally. (It is equally doubtful that Betty killed him unintentionally either.) Or suppose that Reggie wins the lottery, and having bizarre illusions about his ability to control which ticket will win, he enters the lottery and wins it with the intention of winning it [Mele 1997]. The first example suggests that there needs to be some condition added to (6') that says the agent succeeded in G'ing in a manner sufficiently in accordance with whatever plan she had for G'ing as she acted. The second suggests that the agent's success in G'ing must result from her competent exercise of the relevant skills, and it must not depend too much on sheer luck, whether the luck has been foreseen or not. Various other examples have prompted additional emendations and qualifications [see Harman 1976].

There are still more fundamental issues about intentions in action and how they are related to intentions directed at the present and the immediate future. In “Actions, Reasons, and Causes,” Davidson seemed to suppose that ascriptions of intention in action reduce to something like the following.

(7*) The agent F'ed, and at that time he had a pro-attitude toward G'ing and believed that by F'ing he would or might promote G'ing, and the pro-attitude in conjunction with the means-end belief caused him to F'ing, and together they caused it 'in the right way.'

(In Davidson's widely used phrase, the pro-attitude and associated means-end belief constitute a primary reason for the agent to F.) In this account of 'acting with an intention' there is, by design, no mention of a distinctive state of intending. Davidson, at the time of this early paper, seemed to favor a reductive treatment of intentions, including intentions for the future, in terms of pro-attitudes, associated beliefs, and other potential mental causes of action. In any case, Davidson's approach to intention in action was distinctly at odds with the view Anscombe had adopted in Intention. She stressed the fact that constructions like (7) and (7*) supply commonsense explanations of why the agent F'd, and she insisted that the explanations in question do not cite the agent's reasons as causes of the action. Thus, she implicitly rejected anything like (7*), the causal analysis of 'acting with a certain intention' that Davidson apparently endorsed. On the other hand, it was less than clear from her discussion how it is that intentions give rise to an alternative mode of action explanation.

Davidson's causal analysis is modified in his later article “Intending” [1980, essay 5]. By the time of this essay, he dropped the view that there is no primitive state of intending. Intentions are now accepted as irreducible, and the category of intentions is distinguished from the broad, diverse category that includes the various pro-attitudes. In particular, he identifies intentions for the future with the agent's all-out judgments (evaluations) of what she is to do. Although there is some lack of clarity about the specific character of these practical 'all-out' judgements, they play an important role in Davidson's overall theory of action, particularly in his striking account of weakness of will [1980, essay 2]. Despite his altered outlook on intentions, however, Davidson does not give up the chief lines of his causal account of intentions in action — of what it is to act with a certain intention. In the modified version,

(7**) The agent's primary reason for G'ing must cause her, in the right way, to intend to G, and her intending to G must itself cause, again in the right way, the agent's particular act of F'ing.

The interpolated, albeit vague, conditions that require causation in 'the right way' are meant to cover well-known counterexamples that depend upon deviant causal chains occurring either in the course of the agent's practical reasoning or in the execution of his intentions. Here is one familiar type of example. A waiter intends to startle his boss by knocking over a stack of glasses in their vicinity, but the imminent prospect of alarming his irascible employer unsettles the waiter so badly that he involuntarily staggers into the stack and knocks the glasses over. Despite the causal role of the waiter's intention to knock over the glass, he doesn't do this intentionally. In this example, where the
deviant causation occurs as part of the performance of the physical behavior itself, we have what is known as ‘primary causal
deviance.’ When the deviant causation occurs on the path between the behavior and its intended further effects — as in the example of Betty and Jughead above — the deviance is said to be ‘secondary.’ There have been many attempts by proponents of a causal analysis of intention in action (‘causalists,’ in the terminology of von Wright 1971) to spell out what ‘the right kind(s)’ of causation might be, but with little agreement about their success [see Bishop 1989, Mele 1997]. Some other causalists, including Davidson, maintain that no armchair analysis of this matter is either possible or required. However, most causalists agree with Davidson's later view that the concept of ‘present directed intention’ is needed in any plausible causal account of intention in action and acting intentionally. It is, after all, the present directed intention that is supposed to guide causally the ongoing activity of the agent [see also Searle 1983].

The simplest version of such an account depends on what Michael Bratman has dubbed “the Simple View.” This is the thesis that proposition (6) above, [The agent F'd intentionally] and, correspondingly, proposition (7) [The agent F'd with the intention of G'ing] entail that, at the time of action, the agent intended to G. Surely, from the causalist point of view, the most natural account of G'ing intentionally is that the action of G'ing is governed by a present directed intention whose content for the agent is, “I am G'ing now.” So the causalist's natural account presupposes the Simple View, but Bratman [1984, 1987] has presented a well-known example to show that the Simple View is false. He describes a type of case in which the agent wants either to φ or to Θ, without having any significant preference between the two alternatives. The agent does know, however, that it is flatly impossible, in the given circumstances, for him to both φ and Θ although, in these same circumstances, it is open to him to try to φ and try to Θ concurrently. (Perhaps, in trying to φ, he does something with one hand, and, in trying to Θ, he does something with the other.) Believing that such a two-pronged strategy of trying to achieve each goal maximizes his chances of achieving his actual goal of either φ'ing or Θ'ing, the agent actively aims at both of the subordinate ends, trying to accomplish one or the other. The example can be spelled out in such a way that it seems clear that the agent is wholly rational, in his actions and attitudes, as he knowingly pursues this bifurcated attack on his disjunctive goal (but see Yaffe 2010 for skepticism about this claim).

Suppose now that the agent actually succeeds in, say, φ'ing and that he succeeds in virtue of his skill and insight, and not through some silly accident. So, the agent φ's intentionally. It follows from the Simple View that the agent intended to φ. And yet, the agent was also doing something with the intention of Θ'ing and had this attempt succeeded instead (without the intervention of too much luck), then the agent would have Θ'd intentionally. By a second application of the Simple View, it follows that he also intended to Θ. And yet, just as it is irrational to intend to φ while believing that it is flatly impossible for him to φ, so also does it seem irrational to have an intention to φ and an intention to Θ, while believing that it is flatly impossible to do the two things together. So the agent here should be open to criticisms of irrationality in his endeavor to φ or Θ. Nevertheless, we observed at the outset that he is not. The only way out is to block the conclusion that, in trying to φ and trying to Θ in these circumstances, the agent has the contextually irrational pair of intentions, and rejecting the Simple View is the most direct manner of blocking that conclusion.

Even if Bratman's argument defeats the Simple View [see McCann 1986, Knobe 2006], it doesn't rule out some type of causal analysis of acting intentionally; it doesn't even rule out such an analysis that takes the crucial controlling cause to be an intention in every instance. One might suppose, for example, that (i) in a Bratman case, the agent merely intends to try to φ and intends to try to Θ, and that (ii) it is these intentions that drive the agent's actions [Mele 1997]. The analysis in (7**) would be modified accordingly. However, the project of finding a workable and non-circular emendation of (7**) remains an open question.

The conceptual situation is complicated by the fact that Bratman holds that (7) [The agent F'd with the intention of G'ing] is ambiguous between

The agent F'd with the aim or goal of G'ing

and

The agent F'd as part of a plan that incorporated an intention to G.

(8) above is an especially clear example in which the second reading is required. The second reading does entail that the agent intends to F, and it is only the first that, according to Bratman's argument, does not. Therefore, Bratman thinks that we need to distinguish intention as an aim or goal of actions and intention as a distinctive state of commitment to future action, a state that results from and subsequently constrains our practical endeavors as planning agents. It can be rational to aim at a pair of ends one knows to be jointly unrealizable, because aiming at both may be the best way to realize one or the other. However, it is not rational to plan on accomplishing both of two objectives, known to be incompatible, since intentions that figure in rational planning should agglomerate, i.e., should fit together in a coherent larger plan. Bratman's example and the various critical discussions of it have stimulated interest in the idea of the rationality of intentions, measured against the backdrop of the agent's beliefs and suppositions. We discuss some of these issues at greater length in Section 4.

It has been mentioned earlier that Davidson came to identify intentions for the future with all out judgements about what the agent is to be doing now or should do in the relevant future. Velleman [1989], by contrast, identifies an intention with the agent's spontaneous belief, derived from practical reflection, which says that he is presently doing a certain act (or that he will do such an act in the future), and that his act is (or will be) performed precisely as a consequence of his acceptance of this self-referential belief. Paul Grice [1971] favored a closely related view in which intention consists in the agent's willing that certain results ensue, combined with the belief that they will ensue as a consequence of the particular willing in question. Hector-Neri Castañeda [1975], influenced by Sellars [1966] maintained that intentions are
George Wilson [1989] and Carl Ginet [1990] follow Anscombe in holding that reason explanations are distinctively not 'efficient' or 'Humean' or 'producing' causes of action. Some sense of 'reason,' and the statement, "John was caused to seek revenge by his anger," may mean nothing more than, "John's someone says that John has cause to be offended by Jane's truculent behavior, then "cause" in this setting just means circumstances of the debate were not improved by the loose behavior of the ordinary concept of 'a cause.' When the position that reason explanations are causal explanations — explanations that cite the agent's desires, intentions, actions in terms of the agent's reasons for acting. As stated previously, Davidson and other action theorists defended the endorsement of an appropriate type of structured content. When a person believes that $P$, she endorses or accepts the proposition that $P$; when a person intends to $F$, she endorses or accepts the practice, '[I am] to $F$.' Roughly a practition ascribes an action property $F$ to an agent, but the ascription involves a distinctive type of predication that essentially carries some kind of imperative force. Orders, commands, and requests all have practitions as their contents as well, but, as a rule, these will represent prescriptions directed at others. They express the content, e.g., 'You [are] to $F$.' An intention is, by contrast, self-directed, but it is not only that the intended practition is self-directed in this sense; in intending the agent conceives of himself under a distinctively 'first person' conception. Other philosophers, e.g., Hare [1971] and Kenny [1973] have likened intentions to self-directed commands. Still others, notably Annette Bair [1970], have wanted to construe the logical objects of intending as non-propositional and as represented by an unmodified infinitive. Versions of both of these ideas are worked out more carefully and extensively in some of Castañeda's key writings on action. Castaneda was concerned to assign a systematic semantics to the chief locutions that figure in practical thinking and reasoning. These include ascriptions of belief and ascriptions of intention, but they also include the varieties of 'ought' statements that make explicit the normative character of practical reflection. It was a chief ambition in his investigations to chart out the structure of implicational relations that hold between propositions and practitions of these varied sorts and thereby to elaborate the conceptual foundations of deontic logic. (For a rich exegesis of Castañeda on action see essay 12 in Bratman 1999.) Individuals do not always act alone. They may also share intentions and act in concert. There has been growing interest in the philosophy of action about how shared intention and action should be understood. A central concern is whether the sharing of intentions should be given a reductive account in terms of individual agency (see Searle 1990 for an important early discussion of the issue). Michael Bratman [1992] offers an influential proposal in a reductive vein that makes use of his planning conception of intentions. A central condition in his account of shared cooperative activity is that each participant individually intends the activity and pursues it in accordance with plans and subplans that do not conflict with those of the other participants. But Margaret Gilbert [2000] has objected that reductive approaches overlook the mutual obligations between participants essential to shared activity: each participant is obligated to the others to do his or her share of the activity, and unilateral withdrawal constitutes a violation of this obligation. Gilbert argues that a satisfactory account of these mutual obligations requires that we give up reductive individualist accounts of shared activity and posit a primitive notion of joint commitment (see also Tuomela, 2003). Roth [2004] takes seriously the mutual obligations identified by Gilbert, and offers an account that, while non-reductive, nevertheless invokes a conception of intention and commitment that in some respects is friendlier to that invoked by Bratman. It is not entirely clear whether, in positing primitive joint commitments, Gilbert means to commit herself to the ontological thesis that there exist group agents over and above the constituent individual agents. Pettit [2003] defends just such a thesis. He argues that rational group action often involves the "collectivizing of reason," with participants acting in ways that are not rationally recommended from the participant's individual point of view. The resulting discontinuity between individual and collective perspectives suggests, on his view, that groups can be rational, intentional agents distinct from their members.

3. The Explanation of Action

For many years, the most intensely debated topic in the philosophy of action concerned the explanation of intentional actions in terms of the agent's reasons for acting. As stated previously, Davidson and other action theorists defended the position that reason explanations are causal explanations — explanations that cite the agent's desires, intentions, and means-end beliefs as causes of the action [see Goldman 1970]. These causalists about the explanation of action were reacting against a neo-Wittgensteinian outlook that claimed otherwise. In retrospect, the very terms in which the debate was conducted were flawed. First, for the most part, the non-causalist position relied chiefly on negative arguments that purported to show that, for conceptual reasons, motivating reasons could not be causes of action. Davidson did a great deal to rebut these arguments. It was difficult, moreover, to find a reasonably clear account of what sort of non-causal explanation the neo-Wittgensteinians had in mind. Charles Taylor, in his book The Explanation of Action [1964], wound up claiming that reason explanations are grounded in a kind of 'non-causal bringing about,' but neither Taylor nor anyone else ever explained how any bringing about of an event could fail to be causal. Second, the circumstances of the debate were not improved by the loose behavior of the ordinary concept of 'a cause.' When someone says that John has cause to be offended by Jane's truculent behavior, then "cause" in this setting just means 'reason,' and the statement, "John was caused to seek revenge by his anger," may mean nothing more than, "John's anger was among the reasons for which he sought revenge." If so, then presumably no one denies that reasons are in some sense causes. In the pertinent literature, it has been common to fall back on the qualified claim that reasons are not 'efficient' or 'Humean' or 'producing' causes of action. Unfortunately, the import of these qualifications has been less than perspicuous.

George Wilson [1989] and Carl Ginet [1990] follow Anscombe in holding that reason explanations are distinctively
grounded in an agent's intentions in action. Both authors hold that ascriptions of intention in action have the force of propositions that say of a particular act of F-ing that it was intended by its agent to G (by means of F-ing), and they claim that such de re propositions constitute non-causal reason explanations of why the agent F-ed on the designated occasion. Wilson goes beyond Ginet in claiming that statements of intention in action have the meaning of

(9) The agent's act of F-ing was directed by him at [the objective] of G-ing,

In this analyzed form, the teleological character of ascriptions of intention in action is made explicit. Given the goal-directed nature of action, one can provide a familiar kind of teleological explanation of the relevant behavior by mentioning a goal or purpose of the behavior for the agent at the time, and this is the information (9) conveys. Or, alternatively, when a speaker explains that

(10) The agent F-d because he wanted to G,

the agent's desire to G is cited in the explanation, not as a cause of the F-ing, but rather as indicating a desired goal or end at which the act of F-ing came to be directed.

Most causalists will allow that reason explanations of action are teleological but contend that teleological explanations in terms of goals — purposive explanations in other words — are themselves analyzable as causal explanations in which the agent's primary reason(s) for F-ing are specified as guiding causes of the act of F-ing. Therefore, just as there are causalist analyses of what it is to do something intentionally, so there are similar counterpart analyses of teleological explanations of goal directed and, more narrowly, intentional action. The causalist about teleological explanation maintains that the goal of the behavior for the agent just is a goal the agent had at the time, one that caused the behavior and, of course, one that caused it in the right way [for criticism, see Sehon 1998, 2005].

It has not been easy to see how these disagreements are to be adjudicated. The claim that purposive explanations do or do not reduce to suitable counterpart causal explanations is surprisingly elusive. It is not clear, in the first place, what it is for one form of explanation to reduce to another. Moreover, as indicated above, Davidson himself has insisted that it is not possible to give an explicit, reductive account of what 'the right kind of causing' is supposed to be and that none is needed. Naturally, he may simply be right about this, but others have felt that causalism about reason explanations is illicitly protected by endemic fuzziness in the concept of ‘causation of the right kind.’ Some causalists who otherwise agree with Davidson have accepted the demand for a more detailed and explicit account, and some of the proposed accounts get extremely complicated. Without better agreement about the concept of ‘cause’ itself, the prospects for a resolution of the debate do not appear cheerful. Finally, Abraham Roth (2000) has pointed out that reasons explanations might both be irreducibly teleological and also cite primary reasons as efficient causes at the same time. It is arguable that similar explanations, having both causal and teleological force, figure already in specifically homeostatic (feedback) explanations of certain biological phenomena. When we explain that the organism Ved because it needed W, we may well be explaining both that the goal of the V-ing was to satisfy the need for W and that it was the need for W that triggered the V-ing.

In a recent article, Brian McLaughlin (2012) agrees that reason explanations are teleological, explaining an action in terms of a purpose, goal or aim for which it was performed. He also agrees that these purposive explanations are not species of causal explanation. However, he rejects the view that these same explanations are grounded on claims about the agent's intentions in acting, and he thereby sets aside the issues, sketched above, about purpose, intention, and their role in rationalizations. McLaughlin takes the following position: if (i) an agent F-ed for the purpose of G-ing, then, (ii) in F-ing, the agent was thereby trying to G. To assert (i) is to offer an explanation of the action (the F-ing) in terms of the agent's trying to G. Moreover, if (i) is true then the act of F-ing is identical with or is a proper part of the agent's attempt to G. Thus, statement (ii) offers what purports to be, in effect, a mere redescription of the act of F-ing. Assuming Hume's maxim that if an event E causes an event F, then E and F must be wholly distinct, McLaughlin maintains that purposive explanations of actions are constitutive and not causal in character.

Michael Thompson has defended a position that makes a rather radical break from the familiar post-Davidson views on the explanation of action. He rejects as misconceived the debates between causalist and non-causalist accounts of explaining action. He does not deny that actions are sometimes explained by appeal to wants, intentions, and attempts, but he thinks that the nature of these explanations is radically misunderstood in standard theorizing. He thinks that desires, intentions, and their role in rationalizations are not 'propositional attitudes,' as they are usually understood, and the 'sophisticated' explanations that appeal to them are secondary to and conceptually parasitic upon what he calls 'naïve action explanations.' The naïve explanations are given in statements in which one action is explained by mentioning another, e.g., "I am breaking an egg because I am making an omelet." It is a part of the force of these explanations that the explanandum (the egg breaking) is present as part of a broader, unfolding action or activity (the explanans: the omelet making). Similarly, "I am breaking a egg because I am trying to make an omelet," the explanans (the trying) is itself an action, under a certain description, that incorporates the breaking of the egg. Kindred forms such as 'A is F-ing because he wants to G' and 'A is F-ing because he intends to G' are held to give explanations that fall in 'the same categorical space' as the naïve action explanations. Thompson's overall position is novel, complex, and highly nuanced. It is sometimes elusive, and it is certainly not easy to summarize briefly. Nevertheless, it is a recent approach that has rapidly been drawing growing interest and support.

One of the principal arguments that was used to show that reason explanations of action could not be causal was the following. If the agent's explaining reasons R were among the causes of his action A, then there must be some
universal causal law which nomologically links the psychological factors in R (together with other relevant conditions) to the A-type action that they rationalize. However, it was argued, there simply are no such psychological laws; there are no strict laws and co-ordinate conditions that ensure that a suitable action will be the invariant product of the combined presence of pertinent pro-attitudes, beliefs, and other psychological states. Therefore, reasons can't be causes. In "Actions, Reasons, and, Causes," Davidson first pointed out that the thesis that there are no reason-to-action laws is crucially ambiguous between a stronger and a weaker reading, and he observes that it is the stronger version that is required for the non-causalist conclusion. The weaker reading says that there are no reason-to-action laws in which the antecedent is formulated in terms of the 'belief/desire/intention' vocabulary of commonsense psychology and the consequent is stated in terms of goal directed and intentional action. Davidson accepted that the thesis, on this reading, is correct, and he has continued to accept it ever since. The stronger reading says that there are no reason-to-action laws in any guise, including laws in which the psychological states and events are re-described in narrowly physical terms and the actions are re-described as bare movement. Davidson affirms that there are laws of this second variety, whether we have discovered them or not.[6]

Many have felt that this position only lands Davidson (qua causalist) in deeper trouble. It is not simply that we suppose that states of having certain pro-attitudes and of having corresponding means-end beliefs are among the causes of our actions. We suppose further that the agent did what he did because the having of the pro-attitude and belief were states with (respectively) a conative and a cognitive nature, and even more importantly, they are psychological states with certain propositional contents. The specific character of the causation of the action depended crucially on the fact that these psychological states had 'the direction of fit' and the propositional contents that they did. The agent Ped at a given time, we think, because, at that time, he had a desire that represented Fing, and not some other act, as worthwhile or otherwise attractive to him.

Fred Dretske [1988] gave a famous example in this connection. When the soprano's singing of the aria shatters the glass, it will have been facts about the acoustic properties of the singing that were relevant to the breaking. The breaking does not depend upon the fact that she was singing lyrics and that those lyrics expressed such-and-such a content. We therefore expect that it will be the acoustic properties, and not the 'content' properties that figure in the pertinent explanatory laws. In the case of action, by contrast, we believe that the contents of the agent's attitudes are causally relevant to behavior. The contents of the agent's desires and beliefs not only help justify the action that is performed but, according to causalists at least, they play a causal role in determining the actions the agent was motivated to attempt. It has been difficult to see how Davidson, rejecting laws of mental content as he does, is in any position to accommodate the intuitive counterfactual dependence of action on the content of the agent's motivating reasons. His theory seems to offer no explication whatsoever of the fundamental role of mental content in reason explanations. Nevertheless, it should be admitted that no one really has a very good theory of how mental content plays its role. An enormous amount of research has been conducted to explicate what it is for propositional attitudes, realized as states of the nervous system, to express propositional contents at all. Without some better consensus on this enormous topic, we are not likely to get far on the question of mental causation, and solid progress on the attribution of content may still leave it murky how the contents of attitudes can be among the causal factors that produce behavior.

In a fairly early phase of the debate over the causal status of reasons for action, Norman Malcolm [1968] and Charles Taylor [1964] defended the thesis that ordinary reason explanations stand in potential rivalry with the explanations of human and animal behavior the neural sciences can be expected to provide. More recently, Jaegwon Kim [1989] has revived this issue in a more general way, seeing the two modes of explanation as joint instances of a Principle of Explanatory Exclusion. That Principle tells us that, if there exist two 'complete' and 'independent' explanations of the same event or phenomenon, then one or the other of these alternative explanations must be wrong. Influenced by Davidson, many philosophers reject more than just reason-to-action laws. They believe, more generally, that there are no laws that connect the reason-giving attitudes with any material states, events, and processes, under purely physical descriptions. As a consequence, commonsense psychology is not strictly reducible to the neural sciences, and this means that reason explanations of action and corresponding neural explanations are, in the intended sense, 'independent' of one another. But detailed causal explanations of behavior in terms of neural factors should also be, again in the intended sense, 'complete.' Hence, Explanatory Exclusion affirms that either the reason explanations or the prospective neural explanations must be abandoned as incorrect. Since we are not likely to renegade upon our best, most worked-out scientific accounts, it is the ultimate viability of the reason explanations from commonsense 'vernacular' psychology that appear to be threatened. The issues here are complicated and controversial — particularly issues about the proper understanding of 'theoretical reduction.' However, if Explanatory Exclusion applies to reason explanations of action, construed as causal, we have a very general incentive for searching for a workable philosophical account of reason explanations that construes them as non-causal. Just as certain function explanations in biology may not reduce to, but also certainly do not compete with, related causal explanations in molecular biology, so also non-causal reason explanations could be expected to co-exist with neural analyses of the causes of behavior.

4. Intentions and Rationality

Earlier we introduced the Cognitivist view that intentions are special kinds of beliefs, and that, consequently, practical reasoning is a special form of theoretical reasoning. Some theorists of action have been attracted to Cognitivism because of its promise to vindicate Anscombe's (admittedly controversial) claim that, in acting intentionally, we have knowledge of what we are doing that we do not get by observation. But an opposing tradition has been at least as equally prominent in the last twenty-five years of thinking about the nature of intention. Philosophers in this tradition
have turned their attention to the project of giving an account of intention that captures the fact that intentions are distinctive mental states, states which play unique roles in psychological explanations and which are subject to their own sorts of normative requirements.

This project of articulating the distinctive nature of intention was influentially undertaken in Michael Bratman's *Intention, Plans, and Practical Reason* (1987), partially as a response to the reductive view, which had once been endorsed by the early Davidson, according to which intentions could be analyzed as complexes of beliefs and desires. Much contemporary work on normativity and moral psychology can be seen as flowing from Bratman's central (purported) insight about the distinctive nature of intention.

On the simple desire-belief model, an intention is a combination of desire-belief states, and an action is intentional in virtue of standing in the appropriate relation to these simpler states. For example, to say that someone intentionally turns on the air conditioner is just to explain her action by appealing to (e.g.) a desire to turn on the air conditioner and a belief that moving her hand in a certain way is a token of that type of act. It is important to note that Bratman's early arguments were directed against this simple desire-belief model of intention, and not necessarily against the model proposed by Cognitivists. We turn in a moment to the question of to what degree Bratman's theory of intention militates against the latter view.

Bratman motivated the idea that intentions are psychologically real and not reducible to desire-belief complexes by observing that they are motivationally distinctive, and subject to their own unique standards of rational appraisal. First, he noted that intentions involve characteristic kinds of motivational commitment. Intentions are conduct controlling, in the sense that if you intend to \( F \) at \( t \), and nothing changes before \( t \), then (other things equal) you will \( F \). The same is clearly not true for desire; we habitually resist present-directed desires. Second, he noted that intentions involve characteristic kinds of normative commitment (or "reasoning-centered commitment"). Intentions resist reconsideration—they are relatively stable, in the sense that we take ourselves to be settled on a course of action when we intend it, and it seems to be irrational to reconsider an intention absent specific reason for doing so. In addition, intentions put pressure on us to form further intentions in order to more efficiently coordinate our actions. When we intend to go to the park, for example, we feel pressure to form intentions concerning how to get there, what to bring, etc. Again, desires do not appear to be subject to norms of non-reconsideration, and they do not seem to put pressure on us to form further desires about means.

Bratman went on to provide a more rigorous characterization of the constitutive norms on intention, a characterization that has been hugely influential. The three main norms he discussed are requirements of internal consistency, means-end coherence, and consistency with the agent's beliefs. The applicability of these requirements to states of intention was, for Bratman, a further strike against the desire-belief model.

The first norm requires agents to make their intentions consistent with one another. Imagine that Mike intends to go to the game, and also intends to refrain from going. Mike seems obviously irrational. Yet it would be in no way irrational for Mike to desire to go to the game and to desire to refrain from going. So it appears that the irrationality of having inconsistent intentions cannot be explained by appealing to run of the mill norms on desire and belief. Likewise, intentions seem subject to a norm of means-end coherence. If Mike intends to go to the game, and believes that he must buy a ticket in advance in order to go, then he is obviously irrational if he does not intend to buy a ticket (provided he persists in intending to go to the game). Again, merely desiring to go to the game, and believing that going to the game requires buying a ticket, would not be sufficient to render Mike irrational in the event that he failed to desire to buy one. So again it appears that the norms on beliefs and desires cannot suffice to generate the norms on intentions.

Finally, Bratman claimed that rational agents have intentions that are consistent with their beliefs. The exact nature of this intention-belief consistency norm has since been the subject of considerable attention [Bratman 1987, Wallace 2001, Yaffe 2010]. But the general idea is that it is irrational to intend to \( F \) while also believing that one will not \( F \)—this would amount to an objectionable form of inconsistency. Yet desiring to \( F \) while believing that one will not \( F \) seems like no rational error at all.

[I should be noted that the general intuition about the irrationality of this form of inconsistency is by no means unassailable. As Bratman himself points out, it seems perfectly possible, and not irrational, to intend to stop at the library without believing that I will (recognizing, say, my own forgetful nature). If that is correct, then it is not immediately obvious why I could not permissibly intend to stop while also believing that I will not.]

However, while Bratman's arguments do seem devastating for the desire-belief view of intention, they are not necessarily as persuasive against the Cognitivist's reduction of intentions to beliefs. For example, consider again the norm of intention consistency, which convicts Mike of error when he intends to go to the game and also intends to refrain from going. Above we suggested that this norm could not be explained by appealing to norms on desire, since it is permissible to have inconsistent desires. But now imagine that the intention to \( F \) just is (or necessarily involves) the belief that one will \( F \). Then intending to \( F \), and intending to refrain from \( F \)-ing, will entail that one has contradictory beliefs. So if the Cognitivist can help himself to this constitutive claim about the link between intending and believing, he appears to have an attractive explanation of the norm requiring intention consistency. The status of this constitutive claim, and of the plausibility of deriving other norms (e.g. means-end coherence) from it, is a matter of dispute (see Ross 2008). Of course, if Bratman was right to contend that one can intend to \( F \) without believing that she will \( F \), then the Cognitivist picture of intention seems doomed from the get-go.

Seen in another light, then, the conclusion that intentions are psychologically real and irreducible to simpler states may be vindicated by way of a critique of the motivations for Cognitivism. In this vein, some philosophers (notably Sarah


Yaffe, Gideon, 2010, Attempts: In the Philosophy of Action and the Criminal Law, New York, Oxford University Press.

Academic Tools

How to cite this entry.

Preview the PDF version of this entry at the Friends of the SEP Society.

Look up this entry topic at the Indiana Philosophy Ontology Project (InPhO).

Enhanced bibliography for this entry at PhilPapers, with links to its database.

Other Internet Resources

- Action Theory page (Andrei Buckareff, University of Rochester).
- Action Theory (Élisabeth Pacherie, Institut Jean-Nicod, CNRS).

Related Entries

agency | agency: shared | Davidson, Donald | intention | practical reason | reasons for action: justification, motivation, explanation | self-knowledge

Copyright © 2012 by
George Wilson
Samuel Shpall

Open access to the SEP is made possible by a world-wide funding initiative. The Encyclopedia Now Needs Your Support. Please Read How You Can Help Keep the Encyclopedia Free.
ACTION is a bus operator in Canberra, Australia owned by the ACT Government. On 19 July 1926, the Federal Capital Commission commenced operating public bus services between Eastlake (now Kingston) in the south and Ainslie in the north. The service was first known as Canberra City Omnibus Service, but it has had a number of names over the years, including Canberra City Bus Service, Canberra Omnibus Service and Canberra Bus Service. On 14 February 1977, it was renamed as the Australian Capital Territory.