

Action

Luca Ferrero

UW-Milwaukee
Philosophy
ferrero@uwm.edu
www.uwm.edu/~ferrero

to appear in

J.Shand (ed.), *Contemporary Issues in Philosophy*, Blackwell

Version 1.2
May, 2008

This is a draft – subjected to minor corrections prior to publication
Please quote from published version when available

Actions, Agents, and Agency

We are agents. Not only are we capable of acting, but considerable portions of our lives are taken up by our doings, by exercises of our agency. Our actions and doings are essential to much of what we cherish most in our lives, and—arguably—our death can be equated with the permanent loss of our agency. Under these respects, we differ from inanimate objects, artifacts, chemical substances, and natural phenomena such as—for instance—planets, tables, acids, and lightning-storms. When we speak of the ‘actions’ of these things, we simply refer to the operations of some of their characteristic causal powers. But when speak of our own actions and doings, we refer to phenomena that, at least in their paradigmatic form, have all of the following: they are directed at some aim or purpose, they are the subject matter of practical deliberation, the objects of our intentions, the chief manifestations of our freedom, the primary targets of accountability and moral evaluation, and the characteristic objects of demands for rational intelligibility and justification.

What is the nature of our agency and its characteristic manifestations? One might begin addressing this question by considering what difference there is between what *we do* and what merely *happens to us*, between what we perform and what we suffer or undergo. It is uncontroversial that an action is directed at a goal, brings about some transformation in the world, and originates in its agent, at least in the sense that there is a subject who exercises some privileged, direct, and immediate control—although possibly not an exclusive or an ultimate one—over the action’s inception and execution. Despite the familiar and uncontroversial character of these observations, it is far from trivial to

formulate a satisfactory philosophical account of action and agency.

To begin with, the distinctive features of an action might not be externally observable. For instance, when I raise my arm, I make a movement that might be indistinguishable from the one produced, say, by a spasm, i.e., by the mere rising of my arm. Possibilities of this kind suggest that there is more to action than bodily movements alone. Action is, at least in part, a matter of the operation and existence of mental events and states, such as decisions, intentions, desires, and beliefs. Several questions arise at this point: Which psychological features are required for agency? Which mental states and events are required for action? How are these mental components related to each other and to the bodily movement? Is the relation causal, justificatory, or both? If the latter, how are justification and causation related? And what is an action, exactly? The movement produced by the proper mental components, the operation of the mental elements alone, or some combination of the movement and the mental elements? Finally, how do the answers to these questions account for the role of the *agent* and the special importance that we attribute to actions both as expressions of our ‘true selves’ and as proper objects of accountability and responsibility?

A further complication is that agency comes in various kinds and degrees. The philosophy of action is primarily interested in *full-blooded intentional agency*. That is, the agency paradigmatically instantiated by situations where the agent is aware of what she is doing and of why she is doing it, she acts as a result of an explicit deliberation, and she sees her conduct as ‘up to’ her rather than as the product of ‘alien forces’. These cases are distinctive of the agency of adult human beings. The nature of full-blooded intentional agency can be fully understood, however, only by appreciating how it differs

not just from utterly passive happenings but also from lesser kinds of agency—in the spectrum that goes from the complex intelligent purposive behaviors of higher animals to the simpler teleological processes of lower organisms (such as plant phototropism—e.g., the sunflower’s tracking of sunlight—or bacterial chemotaxis—e.g., the movement of a bacterium in response to changes in the gradient of glucose in its surroundings).

Moreover, many things that we describe as our ‘doings’, even as our intentional doings, might fall short of the paradigm of full-blooded intentional agency. For instance, normally we can be said intentionally to fall asleep only in the sense that we intentionally create conditions—say, taking a sleeping pill—that induce us to passively *fall* asleep. The voluntary control of our physiological processes is normally only of this indirect sort. Hence, in their normal operation they are not actions of ours, even if they are the ‘doing’ of our own body (e.g., we can intentionally increase our heartbeat only *by* engaging in some strenuous physical activity or by taking a stimulant). Consider then cases such as sneezing, coughing, and breathing. Although we might have a certain amount of control in inhibiting or *delaying* their occurrences, normally we do not voluntarily initiate them and often we are ultimately unable to resist them. When so, our sneezing, coughing, or breathing are not things that we do in the same intentional way in which we might inhibit or delay their occurrences.

Other cases in which we do not seem to have full intentional control are the behaviors produced by unconscious motivation, compulsion, addiction, and hypnosis. Although we acknowledge that these behaviors originate *within* us, we are reluctant to qualify them as fully intentional since they appear to stem from parts of us from which we are ‘alienated’, i.e., from parts that we do not acknowledge as belonging to our ‘true’ or ‘deep self’.

Finally, an intentional action is not necessarily a deliberate one. It need not be preceded by an explicit and fully articulate deliberation. However, it is the sort of conduct that is the standard subject matter of deliberation. In addition, an intentional action might be executed ‘automatically’, at least in the sense that it might take place outside of the agent’s focus of attention. There is, nonetheless, a point past which automatic execution and lack of awareness of one’s conduct disqualify a purposive behavior from counting as fully intentional, making it more akin to a manifestation of the lesser kind of agency characteristic of non-human animals.

In keeping with the standard focus of the philosophical investigation of agency, this chapter is primarily concerned with the paradigmatic instances of full-blooded intentional agency; with actions such as—to use a standard example—the deliberate flipping of a switch in order to illuminate a room. The focus on scenarios of this kind should not be interpreted, however, as implying that intentional agency necessarily involves bodily movements, brings about positive changes, and is exercised by single agents in isolation. But limitations of space prevent me from discussing the issues raised by mental acts, omissions, and collective agency, respectively. A final introductory remark: the ultimate aim of the philosophical investigation of action is to understand the nature of *agency*, of the capacity that makes us agents and that usually, although not necessarily, is manifested in our actions. Hence, although it is customary to refer to this investigation as ‘philosophy of action’ (and sometimes as ‘action-theory’), ‘philosophy of agency’ would be a better and more comprehensive label for it.

The Explanation of Action

A distinctive feature of actions, as opposed to mere happenings, is that we explain their

occurrences by appealing to reasons rather than mere efficient causes. Elizabeth Anscombe argues that intentional actions are those to which ‘a particular sense of the question ‘Why?’ is given application’, the sense in which ‘the answer, if positive, gives a reason for acting’. An action is the kind of happening that can be made intelligible, rationalized, assessed, and justified by appealing to reasons for it.¹

These reasons need not be explicitly entertained and articulated by the agent. Nor does an agent necessarily act on good reasons. The important point is rather that a conduct is intentional only when it is in principle subjected to a demand for justification in terms of the agent’s reasons for it (a request that in the limiting case might be discharged by claiming that one acted ‘for no reason’).

According to Anscombe, reasons are not efficient causes. Donald Davidson rejects Anscombe’s anti-causalism. He claims that an action is both caused and rationalized by the joint operation of a belief and a desire. For instance, my desire to illuminate the room and my belief that I have the ability and opportunity to illuminate it by flipping the

¹ Anscombe (1963). The beginning of the contemporary philosophical investigation on agency (and of the ‘philosophy of action’ as a distinct area of philosophy) could be dated back to the publication of Anscombe (1963, first edition 1957) and Davidson (1963/1980: Ch.1). Although the nature of action and agency has been the object of philosophical interest at least since Socrates, this investigation was usually pursued only as preparatory to discussing issues in other areas of philosophy—such as the metaphysics of free will, the mind/body problem, and the role of voluntariness and intentionality for moral accountability and evaluation. By contrast, the contemporary discussion has largely proceeded on its own terms. Moreover, only rarely has it engaged with the specific accounts of actions advanced in the history of philosophy (the more notable exception is Korsgaard 2009) even if many contemporary theorists have found some inspiration in the views of action advanced by such diverse group of philosophers as Plato, Aristotle, Aquinas, Hume, Kant, and Wittgenstein. Unfortunately, the philosophy of action in the analytic tradition has for the most part ignored the important works on agency by the early Heidegger, Merleau-Ponty, and Sartre (see especially Sartre 1956), although the views of the former two have recently gain a renewed attention in the discussion of ‘situated and embodied’ cognition in the philosophy of cognitive science.

switch, when properly combined, both cause and rationalize my flipping the switch. These mental states play a dual role. Their efficient causal powers explain the occurrence of the action, their contents rationalize it. Davidson's central argument is that, unless reasons are causes, we cannot account for the distinction between the many reasons the agent might have *to* do something (all the possible justifications she might have for that action) and the reason *for* which she *actually* performed it.²

Thanks to Davidson, a causal account of the nature of action and its explanation in terms of belief/desire pairs (possibly augmented with intentions as distinct mental states) became the new orthodoxy, the 'standard story about action'. This story has not gone unchallenged. As even its defenders noted, so-called deviant causal chains raise serious difficulties. For instance, a climber desires to rid himself of the weight and danger of holding another man on a rope, and believes that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and desire might so unnerve him as to cause him to loosen his hold, but he does not do so intentionally. The belief and the desire do not cause the action 'in the right way'. The problem, which many consider still unresolved, if not unsolvable, is whether an account of 'the right way' can be offered in purely causal terms as the standard story demands.³

George Wilson offers a more radical criticism of the causal nature of action-explanation. According to him, the explanation of action is a species of *non-causal* teleological explanation. Talk of 'the intention with which' a person acted indicates that the act is directed by the agent at a certain objective. It makes explicit the goal-directed

² Davidson (1963/1980: Ch.1).

³ On action explanation, see Melden (1961: Ch.8-9), Davidson (1980: Ch.1; 1987), von Wright (1971: Ch.1), Dretske (1988), Mele (1992).

nature of action but it does not specify one of its causal antecedents. Wilson argues that these teleological explanations cannot be analyzed as causal explanations in which the reasons play the role of guiding efficient causes. The dispute over the nature of action-explanation is still open.⁴

The Standard Story

Throughout the empiricist tradition up to the early twentieth century, the philosophy of agency was dominated by *volitionism*, the view that actions are made intentional and voluntary when caused by distinctive conscious mental occurrences called ‘volitions’ or ‘acts of will’. Gilbert Ryle moved a devastating criticism to classical volitionism by showing that it faces an inescapable dilemma: If volitions are *intentional* acts, they can be made so only by other volitions, which gives rise to an infinite regress; if volitions are mere happenings, instead, it is unclear how the combination of two mere happenings (the volition and the bodily movement caused by the volition) might amount to an intentional action.⁵

The standard story of action, although it appeals to mental states as causal antecedents of action, does not run into Ryle’s dilemma. The belief/desire pair is not the same as an ‘act of will’. Moreover, the standard story offers an informative account of the rationalization of action. The problem with the standard story is rather that it seems to fail to include the agent. For it seems to cast the agent as the mere passive arena for the

⁴ Wilson (1989). See also Schueler (2003). A related debate concerns the issue whether reasons for action should be conceived as (putative) states of affairs rather than mental states. Some argue that the reasons for which the agent flips the switch, say, are things like the ‘world-involving’ fact that, by flipping the switch, she would illuminate the room—not her *belief* that the flipping would work this way, and the fact that there is something desirable about the room’s being illuminated—not her *desire* for it; see Dancy (2000).

⁵ Ryle (1949: Ch.3). For a defense of classical volitionism, see Prichard (1945).

interaction of the mental states that cause the action.

This criticism comes in two forms. Some accept the basic outline of the causal story but argue that we need a more complex picture of the psychology of agents, one that goes beyond simple belief/desire pairs. These views are moved by considerations in so called ‘moral psychology’, they want to account for the agent’s characteristic *identification* with the springs of her full-blooded intentional conduct, as opposed to the alienation she experiences when her conduct does not stem from her true or core self as in the cases of unconscious motivation or compulsion. Other philosophers worry that the standard story leaves the agent out because it misunderstands the nature of the causal relation between agents and actions. They argue that agents are sources of a distinctive kind of causal contribution. This causality does not fit with the strictures imposed by the ‘naturalistic’ reduction of agency attempted by the standard story, a reduction that is usually accepted by those who pursue the first line of criticism. Let’s consider these criticisms in turn.

Agency, Identification, and Reflection

As indicated at the outset, there seems to be an importance difference between merely purposive behaviors (such as those produced by unconscious motivation, compulsion, addiction, and hypnosis) and full-blooded intentional actions. Only the latter ones seem to manifest or stem from the agent’s ‘true self’. Harry Frankfurt accounts for this difference in terms of the notions of ‘guidance’ and ‘identification’. First, goal-directed behaviors are *guided* throughout their temporal unfolding. They are not the simple products of some triggering causal antecedent. They are, rather, sustained by the agent’s ability and readiness to secure the achievement of one’s goal by making compensatory adjustments to her conduct when interfered with, and to stay idle when no adjustment is called for.

Some behaviors are guided by local mechanisms within the agent's body, like the dilating of the pupil in response to the fading of light. Other behaviors are truly guided by the agent since the agent as a *whole* is responsible for the compensatory adjustments.

Guidance by the whole agent is all that is required for the intentional behavior of non-reflective beings, like animals and children, which are inescapably immersed in their purposive conduct. Reflective beings like us, however, might still be *alienated* from the behaviors stem from this global guidance, since they might not identify with their own motives. A drug-addict, for instance, can be guided by her desire for the drug and yet reach for it 'unwillingly' or 'in spite of herself' since she does not *identify* with that motive. Identification is required for the behavior of a reflective agent to be fully intentional.

Frankfurt accounts of identification in terms of hierarchical attitudes of a reflective agent. A reflective agent (as opposed to a non-reflective 'wanton', to use his terminology) has second-order desires about the first-order desires that are to be effective in determining her conduct. An agent identifies with a first-order motive that moves her to act when this is the motive that she desires, at the higher-level, to be effective in moving her to act. The unwilling addict, for instance, is alienated from her conduct because she is moved to reach for the drug by a first-order desire for the drug that goes against her unconflicted second-order desire *not* to be moved by her desire for the drug. Were it not for her addiction—for the irresistibility of the first-order desire, the agent would be expected to prevent that desire from determining her conduct and, if successful, to identify with her refusal to take the drug.⁶

⁶ Frankfurt (1988: Ch. 2, 4-7, 12).

There is a problem, however, with Frankfurt's view. A hierarchy of motives does not appear to account for identification. The fact that a motive is of a higher order does not guarantee that that motive speaks for the agent. A reflective agent might have motives of an even higher order than the second one, motives that in principle could go against the lower-order ones (e.g., the addict might have a third-order desire against the effectiveness of her second-order desire not to be moved by her first-order desire for the drug). The problem is that there seems to be no principled way to determine at which level in this potentially infinite hierarchy of motives we should stop to locate the agent.⁷ Frankfurt's response is centered on the idea of 'satisfaction': an agent identifies with a first-order motive when she has a second-order motive for the effectiveness of the first order and she is *satisfied* with the second-order motive in the sense that she has no active interest in changing it. Notice that the satisfaction is a property of the agent's whole psychic structure. The satisfaction described the *absence* of a pressure for change, not a distinct attitude. If the latter, there would still be a threat of regress, since one could continue raise the question of whether the agent identifies (and it is thus satisfied) at a higher order with her lower-order satisfaction.⁸

David Velleman and Michael Bratman agree with Frankfurt that the standard story of action is wanting and that reflection and hierarchy are fundamental to agential guidance, but they maintain that the notion of 'satisfaction' is inadequate to stop the regress. A depressed, bored, or lazy agent might have no interest in changing her higher-order motives. She would be satisfied in Frankfurt's sense, but this does not appear a case of

⁷ For the criticism of hierarchy and an outline of a non-hierarchical view of identification in terms of the agent's valuing, see Watson (1975).

⁸ Frankfurt (1999: Ch.8).

genuine identification with one's motives since conditions such as depression, boredom, and laziness always carry the potential for reflective dissociation. According to Velleman, the trouble with Frankfurt is his appeal to second-order motives that do not necessarily arise out of an appreciation of the role of first-order motives as *reasons* for action. Velleman argues that being reflective as a rational agent is a matter of being disposed to do what is justified, to do what makes sense to oneself. More precisely, a matter of a 'higher-order motive of rationality' to be moved by a lower-order motive in its capacity as a reason: to acquiesce in being moved by the intrinsic force of a first-order motive only if being so moved is intelligible to the agent (where this intelligibility is a kind of self-knowledge, as explained later). The motive of rationality operates by *reinforcing* pre-existing first-order motives. In full-blooded intentional agency, the agent does not simply flip the switch as a result of the desire to illuminate the room and the belief that flipping it is an effective means (as she would if she were to act impulsively or out of subconscious motives). Rather, her motive is strengthened by the fact that this conduct makes sense to her by comparison to courses of action that, although supported by her first-order motives, are not equally intelligible to her.

There is no regress in Velleman's account because the agent cannot dissociate from the higher-order motive of rationality. This motive drives practical thought and, as such, cannot be made the object of detached critical reflection. The agent is *functionally identical* with the operation of the motive of rationality. She is identical with the capacity for reflection rather than with other specific higher-order desires. Hence, a subject *qua* rational agent cannot ever be alienated from this capacity. She could only disown it by giving up making rational assessments of her motives, i.e., by giving up being a rational

agent.⁹

Bratman agrees with Frankfurt and Velleman's criticisms of the limitations of the simple psychological structure of the standard story. He maintains that full-blooded intentional agency results from the integration of the capacity for reflection with the distinctive diachronic dimension of our temporally extended agency. For him, attitudes 'speak for the agent' only when their role in the subject's psychology partly constitutes and supports her existence as one and the same agent over time. Bratman subscribes to a Lockean theory of personal identity according to which identity is a matter of psychological continuity. An important contribution to this continuity is provided by 'self-governing policies', intention-like attitudes that offer general guidelines about which desires one is to treat as reasons in practical reasoning. When one guides one's thinking and acting in accordance to self-governing policies, one exercises self-governance in one's capacity as an agent because, first, these policies contribute to one's identity over time and thus have authority to speak for oneself and, second, one directs one's thinking and acting in terms of what one takes not as mere motives but as one's reasons for action.

Bratman differs from Velleman in maintaining that identification with a first-order desire is not produced by a *single* higher-order motive or rationality shared by all agents in their capacity as agents. Identification is, rather, due to self-governing policies that can differ from agent to agent. What is common to all agents is only the basic structural role that specific self-governing policies play in securing the temporal identity of each individual agent. Bratman agrees with Frankfurt and Velleman on the importance of

⁹ Velleman (2000: Ch.1, 6).

reflection and its hierarchical structure (self-governing policies are higher-order attitudes about first-order motives). However, he concedes to the critics of hierarchical views that the attitudes that speak for the agent might not be higher-order ones, although he insists that the nature of self-governance puts pressures toward the existence of a hierarchy.¹⁰

Reflection plays a prominent role also in Christine Korsgaard's theory of action. She argues that action is necessarily performed by a unified agent; it is an expression of the agent as a whole rather than a product of forces at work in her. Nonetheless, the agent does not exist as a unified author prior to the action. The agent constitutes herself as such author in the very act of choice: Action is self-constitution. This is true for agency in general, not just for human agency, although different kinds of agents constitute themselves as different forms of life.

Consider *animal* action. According to Korsgaard, the animal is presented with an 'incentive', a motivationally loaded representation of an object. The animal is aware of some features of the object as desirable or aversive in some specific way (e.g., as to be eaten) and she acts on this incentive on the basis of a 'principle', which determines what the animal does in the face of that specific incentive (in the example, to eat the object). The principles on which an animal acts are its instincts. They *automatically* tell the animal which responses are appropriate for each particular incentive. The instinctual operation of the incentives is causal but it does not bypass the animal's own guidance. For the instincts are the laws of the animal's causality; they define the animal's will. By operating on them, the animal is not just purposive but also autonomous, at least in the sense that its movements are determined by its own nature.

¹⁰ Bratman (2007).

The self-determination in *human* action is deeper. We are not simply governed by the principles of our own causality: We choose these principles. We are self-conscious and thus aware of the working of incentives within us. We no longer experience incentives as demands but as proposals. Incentives become ‘inclinations’, something we now have to decide whether to satisfy or not. It is only within this space of ‘reflective distance’ that the question arises whether our incentives give us *reasons* to act. Self-consciousness creates the need for principles of reason.

Animal action is purposive since it is guided by a conception of its object. But our self-conscious action requires a conception of its purpose, of what we are doing and why. In this sense, we are agents who adopt intentions. We are conscious of our own causality and it is thus up to us how we exercise it. Our self-consciousness is the source of a psychic complexity unknown to animals. We are conscious of the threats to our psychic unity. For animals, psychic unity is a natural state: The instincts immediately tell them how to deal with incentives. For us, psychic unity is to be achieved. Being a person is being engaged in a particular form of life: the activity to constitute oneself as a *particular* individual given that, as a *reflective* animal, each of us must create, through one’s own choices, one’s individual self-maintaining form, one’s individual ‘practical identity’.¹¹

Actions and Agents

Let’s now consider the second line of criticism against the standard story of action. This criticism concurs with the first one in denouncing the standard story for leaving the agent out of the picture, but it claims the fault does not lie in overlooking the complexities of reflection but in misunderstanding the nature of the causal relation between agents and

¹¹ Korsgaard (2009).

their actions. Against the standard story, Roderick Chisholm argues that agents, not their mental states, cause actions. Agents do so by being *additional* primitive elements in the explanatory order.¹² The suggestion that there is a distinctive kind of ‘agent-causation’, however, has been widely criticized because it takes agents as *intruders* among natural events in violation with a widely accepted naturalistic conception of causal explanation.¹³

Jennifer Hornsby offers a different criticism of the causal claims of the standard story. She argues that the purely event-based account of the causal order implicit in the standard story misses the agent’s distinctive contribution. The first mistake is to conceive of action as the event of body movement, understood intransitively, rather than as the *moving* of the body, understood transitively. If we conceive of action as a body movement, we are induced to explain it by taking the causes of action to be either the mental states of the standard story (thereby missing the agent’s role), or the agent itself as in agent-causation (thereby making the agent an intruder in the causal order). According to Hornsby, agents cause bodily movements but these movements are not actions, they are only the effects of actions. An action is, rather, the agent’s *causing* of the bodily movements (or, better, her *trying* to move the body, see below) and of the other causal consequences of these movements. In the light-switching scenario, for instance, the action is not the finger movement but the *moving* of the finger, i.e., the agent’s causing of the finger movement. The agent does not cause the moving of the finger, she causes the finger movement and this *causing* is the action. In a similar fashion, the agent is the cause of the other effects of her action, such as the illumination of the room, but her action is not the occurrence of these events, it is her causing them.

¹² Chisholm (1976).

¹³ On agent-causation, see Clarke (1993), Alvarez & Hyman (1998), O'Connor (2000).

Hornsby claims that in explaining an action we are not looking for a causal explanation of the occurrence of a bodily movement or any other effects of the action. We already have an explanation of this occurrence in that we know what caused it, namely, the agent. It is exactly because we already think of this occurrence as the effect of an action that we are interested in understanding *why* the agent caused that effect. In looking to explain the action, we do not want to learn the causal role played by the agent. This is something that we already assume in looking for an explanation of an *action*. We, rather, want to learn things about the agent that make it understandable that she should have brought those effects about, that she should have played the causal role that makes her the author of that action.¹⁴

The Individuation of Action

When I illuminate a room *by* flipping the switch *by* moving my finger, how many actions do I perform? We might be tempted to say that I am doing at least three separate things. Anscombe and Davidson argue, instead, that I am performing only one action, although one that admits of as many descriptions as its disparate causal effects. For them, there is only one event that counts as my action. This event can be described in terms of any of its effects (for instance, as my moving the finger, my flipping the switch, or my illuminating the room).¹⁵

If a causal effect of the illumination of the room (and thus of the finger movement) is that a burglar is alerted of my presence, my action can also be described as my alerting the burglar. This is not to say that by flipping the switch I *intentionally* alert the burglar. Only some of the possible descriptions of the action indicate what I do intentionally. At

¹⁴ Hornsby (2004).

¹⁵ Anscombe (1963); Davidson (1971/1980: Ch.3).

the very least, I am not acting intentionally under any of the descriptions that apply to my action unbeknownst to me. If I have no idea that there is a burglar, I am not intentionally alerting him by flipping the switch, even if the action of unintentionally alerting the burglar is the same action as my intentionally illuminating the room and as my intentionally flipping the switch.¹⁶

We illuminate the room by flipping the switch. We flip the switch by moving our fingers. However, we do not seem to move our fingers *by* doing anything else. The movement of our finger is a 'basic action'. An action is basic when the doing is described in such a way that one cannot be said to be acting under that description *by* doing anything else. (This is not to deny that there are causal antecedents of the basic action within the agent, e.g., muscles contractions and neurophysiologic events, but these are descriptions of what one *does*.) Davidson claims that all basic actions are bodily movements. Whatever we do, we do by moving our bodies, and we do not move our bodies by doing anything else.¹⁷

Contra Davidson, Brian O'Shaughnessy and Jennifer Hornsby argue that a basic action is not a bodily movement but one's *trying* to move one's body. For instance, if someone does not know that her arm is completely paralyzed and she attempts to move her finger, she seems to have done something even if her arm and finger have not moved at all; she has *tried* to move her finger. These philosophers argue that even when we actually succeed in moving our bodies, whatever we do it is something that we achieve *by trying* to do it. The basic description of action is thus always in terms of 'trying', even if in

¹⁶ On the individuation of action, see also Anscombe (1979), Goldman (1970: Ch.1-2), Ginet (1990: Ch.3).

¹⁷ On basic actions, see Danto (1963) and Davidson (1971/1980: Ch.3).

ordinary talk, for pragmatic reasons, we reserve the expression ‘to try’ to describe cases where we either fail to move our bodies or we suspect that it is very likely that we might be unable to move them.¹⁸ This view should not be confused with the classical volitionism presented above. The trying *is* the intentional acting. It is not a distinct phenomenon that accounts for the action’s intentionality and voluntariness as its causal precursor—whence the immunity from Ryle’s criticism of classical volitionism.

Acting Intentionally and Intention

So far I used ‘acting intentionally’ to refer to full-blooded intentional agency. But in ordinary talk we often use ‘intentionally’ more liberally to refer to several distinct, although not necessarily unrelated, aspects of agency. Sometimes we describe a conduct as intentional to indicate that it is goal directed or, more strongly, that it is guided by the agent as a whole rather than a local mechanism. Sometimes we speak of acting intentionally in the narrower sense of acting deliberately, i.e., acting in view of a goal adopted in an explicit and articulate deliberation. In certain contexts, to say that something is done intentionally means that it is not done inadvertently or accidentally. We are also reluctant to claim that something is done intentionally when its outcome, although not deviantly caused, depends to a large extent on chance—e.g., we intentionally roll the dice but we do not intentionally roll a seven with the dice. Doing something intentionally, in this sense, means that one exerts the kind of control and guidance normally expected by proficient agents in the unfolding of that particular activity.¹⁹

¹⁸ On trying, see O’Shaughnessy (1973; 1980) and Hornsby (1980: Ch.1-3).

¹⁹ For a representative sample of the various approaches to the study of intention and intentional action, see Anscombe (1963), Austin (1990: Ch. 8, 12), Davidson

There is an important connection between acting intentionally and the agent's knowledge of what she is doing. Knowledge of what one is doing is necessary for the intentionality of one's action in that one cannot be said to be acting intentionally under a description of the action in terms of unknown and unexpected effects (as in the burglar case previously discussed). But knowledge of the effects is not sufficient to make the action intentional under that description. For instance, if I know that there is a burglar and that my turning on the light is necessarily going to alert him, it might still be that, by intentionally flipping the switch, I do not intentionally alert the burglar; I only *knowingly* do so. Alerting the burglar is an effect of my illuminating the room that I foresee, but I do not intend. Whereas illuminating the room is my goal, alerting the burglar is not. I guide my conduct so as to ensure that it succeeds in illuminating the room. That is, I am expected to make the necessary adjustments to turn on the light. But since I am not aiming at alerting the burglar, I am not making sure that I succeed at it. If it turned out that in illuminating the room I would not be alerting the burglar, I would be under no expectation to find alternate means to alert him.²⁰

According to Anscombe, there is another important relation between intentionality and knowledge: in acting intentionally the agent knows what she is doing 'without observation'. Anscombe's suggestive but somewhat unclear discussion has recently spurred an interesting debate on the relation between self-knowledge and intentional

(1978/1980: Ch.5), Hunter (1978), O'Shaughnessy (1980: Ch.17), Searle (1983: Ch.3), Bratman (1987; 1999; 2007), Velleman (1989; 2000), Mele & Moser (1994), Scheer (1994), and Hartogh (2004).

²⁰ On the difference between intended and merely foreseen effects, see Harman (1986: Ch.9) and Bratman (1987: Ch.10). Notice that the difference between doing something intentionally and doing something knowingly but non-intentionally might make a difference in the assessment of the agent's blameworthiness and culpability, as suggested by the so-called doctrine of 'double effect'.

agency. If the knowledge in question is of the intention as the objective of one's action, this knowledge can be claimed to be non-evidential because it is produced by the agent's first-personal responsiveness to the deliberative considerations that support the adoption of the intention. It does not result from a third-personal investigation about one's mental states on the basis of epistemic grounds, including those provided by introspection.²¹

What about the knowledge of one's actual and future intentional *performance*, the kind of knowledge that seems the primary concern of Anscombe's remarks? George Wilson argues that because of its nature as a practical commitment, an intention comes with the expectation of the stability of the case for its adoption. This expectation provides a defeasible ground for the belief that one will continue to carry it out. But the expectation of a stable case is based not on inductive evidence about the immutability of the agent's preferences but on the agent's continuous sense of the intrinsic force and authority of the reasons for action that have been decisive in his adoption of the intention.

Velleman claims that a different lesson is to be learnt from Anscombe: An intention amounts to knowledge only if it appropriately and reliably causes the facts that make the intention true. For him, intention amount to this knowledge and it does so in a non-evidential fashion: an intention is a *cognitive* commitment to the truth of the intention's content; a commitment that provides a reliable connection to the intended action via the operation of the higher-order motive of rationality (as discussed above). The agent's commitment to the truth of doing what she intends to do reliably enlists reinforcement for the motives that favor that action since that is the only action that satisfies the higher-order motive of rationality. It is the only action that makes true the agent's belief

²¹ Moran (2001).

that she is going to do what she intends to do. The kind of self-knowledge provided by the intention, therefore, counts as ‘practical knowledge’ in Anscombe’s sense, a knowledge that causes what it represents.²²

What is the relation between acting intentionally and having an intention? Anscombe maintains that acting intentionally just consists in one’s conduct being subjected to the demand for explanation in terms of reasons. The term ‘intention’ does not denote a distinct mental state. Defenders of what might be called a ‘behaviorist’ or ‘outward-looking’ conception of agency concur. For them, our talk of intentions indicates distinctive structural features of our performance, including its goal, without committing us to the existence of intentions as distinct causal elements in our psychology. However, one needs not embrace the outward-looking conception to deny the existence of intentions as distinct mental states. For instance, Davidson initially argued that acting intentionally only requires causation by a belief/desire pair. Later he acknowledged the need for intentions as distinct attitudes in order to account for those cases in which one reaches a conclusion about what to do well in advance of the time of action, a conclusion that one might never carry out, as it happens for instance when one is weak-willed. In this later work, Davidson identifies intentions with what he calls ‘all-out value judgments’ about the desirability of an action.²³

According to Bratman, however, a psychology of belief/desires even if augmented with Davidson’s style intentions is too austere to account for the distinctive *planning* structure of our diachronic agency. Bratman argues that prospective intentions are partially specified plans about our future conduct, plans to be filled in as they unfold over

²² Velleman (2000).

²³ Davidson (1980: Ch.1, 5).

time. Intentions *settle* what we are going to do in the future in a way that is usually effective in determining our future conduct. An intention allows the agent to take advantage of more favorable conditions for deliberation in advance of the time of action. It provides a filter for future practical reasoning since one needs only consider options compatible with one's intentions thereby reducing the costs of contingency planning. Last but not least, by settling future conduct, intentions greatly contribute to both intra- and inter-personal coordination of action over time.

For Bratman, intentions are not reducible to a combination of beliefs and desires. They are functionally characterized by a distinctive set of rational pressures. The agent who intends to ϕ is under several rational requirements: She is to be instrumentally coherent (she is either to take the necessary means to ϕ -ing or to abandon her intention); her intention is to be consistent with her beliefs (in particular, she is not to believe that her ϕ -ing is impossible); the intention is to be agglomerated (if the agent intends to ϕ and she intends to ψ , she is also to intend to $\langle\phi$ and $\psi\rangle$); finally, the intention is to be stable over time. By comparison, none of these requirements apply to the agent's desires. Moreover, contrary to the claims of what Bratman calls the 'simple view', when one does something intentionally one does not necessarily act with a corresponding intention. At times, we might be pursuing a goal without trying to fit it within a more complex and global coordinated plan of action. When so, one does not have an intention as a genuinely planning attitude but rather a 'settled objective'—an objective that is under a pressure for instrumental coherence but not for agglomeration. Our truly distinctive form of agency, however, is the planning agency that calls for the global demands for coordination characteristic of full-fledged intentions, rather than the simpler *local* constraints imposed

by ‘settled objectives’ (which, as such, might be the distinctive form of the diachronic agency of the purposive but non-planning agency of non-human animals).²⁴

An important and still debated issue about intentions concerns the source of their effectiveness in controlling future conduct. Do they simply cause the intended future conduct or do they, rather, exert rational authority over it? If the latter, what is the nature and source of its authority? Does an intention generate a reason for the intended action *additional* to the considerations that made the action choiceworthy for the agent when the intention was first adopted? If so, Bratman worries that an intention could have an undesirable ‘bootstrapping effect’: when the time of action comes the agent might find herself with a (possibly decisive) reason to act as intended even if she no longer finds the action choiceworthy independently of her intention to perform it.

Finally, for Bratman an intention is not a cognitive but a practical commitment. Adopting the intention is not to discover something about oneself; it is to make a further practical move, although one that might be accompanied by some knowledge about one’s future conduct. Hence, Bratman rejects cognitivism about practical reason.²⁵ The cognitivists claim that the rational demands for consistency and coherence of intentions are grounded in the norms of *theoretical* rationality, in rational demand for consistency and coherence of belief. This is so because an intention to φ is deemed to be either identical to or to entail a belief that one will φ . A cognitivist like Velleman argues that, if one does not believe that one is going to do what one intends to do, there are no grounds

²⁴ Bratman (1987; 1999).

²⁵ Bratman (2009).

for the coordinating role of intentions and for their distinctive rational pressures.²⁶

Bratman's response in a still ongoing debate is twofold. First, an intention does not necessarily involve a belief about its eventual success. Intention only entails the belief in the possibility of acting as intended. Second, the norms of intention are fundamentally practical. They are grounded not on our nature as cognitive beings but on the requirements for the effectiveness of the distinctive planning character of our diachronic intentional agency.²⁷

²⁶ Velleman (2007).

²⁷ On the relation between intention, prediction, and self-knowledge, see Hampshire (1975: Ch.3), Grice (1971), Velleman (2007), Wilson (2000), Falvey (2000), Moran (2001; 2004). On cognitivism about practical reason, see Harman (1976), Bratman (1999: Ch.13; 2009), Velleman (2000: Ch.1), Wallace (2001), Setiya (2007).

Further Readings [shorter published version]

The best short introduction to the philosophy of action is Wilson (2007). The best book-length introduction is Stout (2005). Three short papers that could serve as introductions to important issues in the philosophy of action are Davidson (1980: Ch.3), Frankfurt (1978), and Kenny (1992: Ch.3). A good collection of essays is Mele (1997).

Further Readings [longer version – not published]

Introductory readings on action and agency: Stout (2005), Wilson (2007), Davidson (1971), Frankfurt (1978), Kenny (1992: Ch.3), Mele (1997).

On **action explanation**: Melden (1961: Ch.8-9), Davidson (1963), Taylor (1964), von Wright (1971: Ch.1), Davis (1984), Davidson (1987), Dretske (1988), Ginet (1989), Mele (1992), Roth (1999), Ruben (2003).

On action and **teleological explanation**: Collins (1987 Ch.4), Wilson (1989), Schueler (2003), Sehon (2005).

On **agency, identification, and reflection**: Frankfurt (1988), Watson (1975), Frankfurt (1992), Velleman (1992), Velleman (2000: Ch.1), Korsgaard (2009), Bratman (2007)

On **agent-causation**: Taylor (1966), Chisholm (1976), Clarke (1993), O'Connor (2000). On the relation between agents and action: Alvarez, Hyman (1998), Hornsby (2004a), Hornsby (2004b).

On the **individuation of action**: Anscombe (1963), Davidson (1967), Goldman (1970: Ch.1-2), Davidson (1971), Anscombe (1979: 219), Ginet (1990: Ch.3).

On **basic actions**: Danto (1963), Danto (1965), Davidson (1971), Danto (1976), Annas (1978)

On classical **volitionism**: Prichard (1945), Ryle (1949 Ch.3). On trying: O'Shaughnessy (1973), O'Shaughnessy (1980), Hornsby (1980: Ch.1-3), Ginet (1990), Pietroski (1998).

On **acting intentionally and intention**: Anscombe (1963), Austin (1956), Austin (1966), Castaneda (1975), Davidson (1978), Hunter (1978), O'Shaughnessy (1980: Ch.17), Searle (1983: Ch.3), Bratman (1987), Velleman (1989), McCann (1991), Mele, Moser (1994), Scheer (1994), Bratman (1999), Velleman (2000), Hartogh (2004), Bratman (2007).

On the difference between **intended and merely foreseen effects**: Harman (1986: Ch.9), Bratman (1987: Ch.10).

On the relation between **intention, prediction, and self-knowledge**: Anscombe (1963), Hampshire (1975: Ch.3), Grice (1971), Velleman (1989), Wilson (2000), Falvey (2000), Moran (2001), Moran (2004), Velleman (2007).

On **cognitivism about practical reason**: Harman (1976), Bratman (1999 Ch.13), Velleman (2000: Ch.1), Wallace (2001), Velleman (2007), Setiya (2007), Bratman (2009).

Other notable works in the philosophy of action: von Wright (1963), Thomson (1977), Bishop (1989), Gustafson (1986), Audi (1993), Bennett (1995), Pink (1996), Rundle (1997), McCann (1998), Juarrero (2002), Mele (2003).

Bibliography

- Alvarez, M. and Hyman, J. (1998) Agents and Their Actions. *Philosophy* 73, 219-245.
- Anscombe, G.E.M. (1963) *Intention*. 2nd edn. Cornell University Press, Ithaca, NY.
- Anscombe, G.E.M. (1979) Under a Description. *Nous* 13, 219-233.
- Austin, J.L. (1990) *Philosophical Papers*. Clarendon Press, Oxford.
- Bratman, M. (1987) *Intentions, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA.
- Bratman, M. (1999) *Faces of Intention*. Cambridge University Press, Cambridge.
- Bratman, M. (2007) *Structures of Agency*. Oxford University Press, Oxford.
- Bratman, M. (2009) Intention, Belief, Practical, Theoretical. In: Timmerman, J., Skorupski, J. & Robertson, S. (eds.) *Spheres of Reason*. Oxford: Oxford University Press.
- Chisholm, R. (1976) The Agent as Cause. In: Brand, M. and Walton, D. (eds.) *Action Theory*. Reidel, Dordrecht, pp. 199-211.
- Clarke, R. (1993) Toward A Credible Agent-Causal Account of Free Will. *Nous* 27, 191-203.
- Dancy, J. (2000) *Practical Reality*. Oxford: Oxford University Press.
- Danto, A.C. (1963) What We Can Do. *Journal of Philosophy* 60, 435-45.
- Danto, A.C. (1965) Basic Actions. *American Philosophical Quarterly* 2, 141-148.
- Davidson, D. (1980) *Essays on Actions and Events*. Clarendon, Oxford.
- Davidson, D. (1987b) Problems in the Explanation of Action. In: Pettit, P., Sylva, R. & Norman, J. (eds.) *Metaphysics and Morality*. Blackwell, Oxford.
- Dretske, F. (1988) *Explaining Behavior*. MIT Press, Cambridge, Mass.
- Falvey, K. (2000) Knowledge in Intention. *Philosophical Studies* 99, 21-44.
- Frankfurt, H.G. (1988) *The Importance of what we care about*. Cambridge University Press, Cambridge.
- Frankfurt, H.G. (1999) *Necessity, Volition, and Love*. Cambridge University Press, Cambridge.

- Ginet, C. (1990) *On Action*. Cambridge University Press, Cambridge.
- Goldman, A. (1970) *A Theory of Human Action*. Prentice-Hall, Englewood Cliffs, NJ.
- Grice, H.P. (1971) *Intentions and Certainty*. *Proceedings of the British Academy* 57, 263-279.
- Hampshire, S. (1975) *Freedom of the Individual*. Expanded edn. Harper, New York.
- Harman, G. (1976) *Practical Reasoning*. *Review of Metaphysics* 79, 431-263.
- Harman, G. (1986) *Change in View: Principles of Reasoning*. MIT Press.
- Hartogh, d.G. (2004) *The Authority of Intention*. *Ethics* 115, 6-34.
- Hornsby, J. (1980) *Actions*. Routledge, London.
- Hornsby, J. (2004) *Agency and Actions*. In: Steward, H. and Hyman, J. (eds.) *Agency and Action*. Cambridge University Press, pp. 1-23.
- Hunter, J.F.M. (1978) *Intending*. Dalhousie University Press, Halifax.
- Kenny, A. (1992) *The Metaphysics of Mind*. Oxford University Press, Oxford.
- Korsgaard, C. (2009) *Self-Constitution: Action, Identity and Integrity*. Oxford: Oxford University Press.
- Melden, A.I. (1961) *Free Action*. Routledge Kegan & Paul, London.
- Mele, A.R. and Moser, P.K. (1994) *Intentional Action*. *Nous* 28, 39-68.
- Mele, A.R. (1992) *The Springs of Action*. Oxford University Press, New York.
- Mele, A.R. (ed.) (1997) *The Philosophy of Action*. Oxford Univ. Press, Oxford.
- Moran, R. (2001) *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press, Princeton, NJ.
- Moran, R. (2004) *Anscombe on 'practical knowledge'*. In: Steward, H. and Hyman, J. (eds.) *Agency and Action*. Cambridge University Press, pp. 43-68.
- O'Connor, T. (2000) *Persons and Causes: The Metaphysics of Free Will*. Oxford University Press, New York.
- O'Shaughnessy, B. (1973) *Trying as the mental 'pineal gland'*. *Journal of Philosophy* 70, 365-386.
- O'Shaughnessy, B. (1980) *The Will*. Cambridge University Press, Cambridge.

- Prichard, H.A. (1945) *Acting, Willing, Desiring*. In: Prichard, H.A. (ed.) *Moral Obligation*. Clarendon Press, Oxford, pp. 187-198.
- Ruben, D.H. (2003) *Action and its Explanation*. Oxford University Press, New York.
- Ryle, G. (1949) *The Concept of Mind*. Chicago: Chicago University Press.
- Sartre, J.P. (1956) *Being and Nothingness*. Translated by Hazel E. Barnes. Philosophical Library, New York.
- Scheer, R.K. (1994) *The causal theory of intentions*. *Philosophical Investigations* 17, 417-434.
- Schueler, G.F. (2003) *Reasons and Purposes: Human Rationality and the Teleological Explanation of Action*. Oxford University Press, Oxford.
- Searle, J.R. (1983) *Intentionality*. Cambridge University Press, Cambridge.
- Setiya, K. (2007) *Reasons without Rationalism*. Princeton University Press, Princeton, NJ.
- Stout, R. (2005) *Action*. McGill-Queen's University Press.
- Velleman, J.D. (1989) *Practical Reflection*. Princeton: Princeton University Press.
- Velleman, J.D. (2000) *The Possibility of Practical Reason*. Oxford University Press, Oxford.
- Velleman, J.D. (2007) *What Good is a Will?* In: Leist, A. and Baumann, H. (eds.) *Action in Context*. De Gruyter/Mouton, Berlin/New York.
- von Wright, G.H. (1963) *Norm and Action*. Routledge, London.
- von Wright, G.H. (1971) *Explanation and Understanding*. Cornell University Press, Ithaca, NY.
- Wallace, R.J. (2001) *Normativity, Commitment, and Instrumental Reason*. *Philosophers' Imprint* 1, 1-26.
- Watson, G. (1975) *Free Agency*. *The Journal of Philosophy* 72, 205-220.
- Wilson, G. (1989) *The Intentionality of Human Action*. Stanford University Press, Palo Alto, CA.
- Wilson, G. (2000) *Proximal Practical Foresight*. *Philosophical Studies* 99, 3-19.
- Wilson, G. (2007) *Action*. In: Zalta, E. (ed.) *Stanford Encyclopedia of Philosophy*.