This volume contains David Velleman’s major papers in action theory and related areas since the publication of his seminal book *Practical Reflection* in 1989. With the exception of the introduction, the essays in this collection all have appeared elsewhere. This is an excellent collection that brings together several important essays. There are many interesting and important ideas in these pages. The introduction and three chapters (6, 7 and 8) form the core of this volume. All deal with the question as to what constitutes agency. In what follows, I concentrate on these papers.

The problem for any plausible theory of action, as Velleman sees it, is to find the necessary and sufficient conditions for determining whether some behavior is an action. The traditional Humean answer to this question is the following: an action is behavior that is caused by the desires and beliefs of the agent (p. 5). However, Velleman thinks it is not enough to distinguish between mere behavior and those behaviors that are caused by the agent. We have to have a theory that enables us to make a second, more important distinction between the agent’s actions and his activities. Velleman gives several examples to illustrate this point. Here is one of mine. Sometimes we say something that we really did not mean to say. For example, my father sometimes mixes up the names of his daughters, addressing my sisters by each other’s names. When this happens, he is not aware that he is making such a mistake. Typically, he is surprised and embarrassed to find out that he did confuse my siblings’ names – again. Such slips are not merely behavior: they are things that the agent does and that are caused by his desires and beliefs. Therefore, they would count as action on the standard Humean story. However, it is clear that they are not instances of real full-blooded action. It is not something that my father intentionally and consciously does. It ‘happens’ to him. That is to say, these slips are not his actions, though they are his activities (p.1-5).

The reason that we intuitively can make a distinction between the activities of the agent and his actions in the full sense is, so claims Velleman, that the agent does not really have a place in
the former. No doubt my father’s mistakes are caused by some brain states that we functionally
can characterize as his desires and beliefs. However, what does he have to do with this? How are
these slips attributable to him? Where does he fit in the story?

This omission in the traditional explanation of action has been recognized before. For
example, Roderick Chisholm argued in the seventies that this omission is the straightforward
result of the naturalism that characterizes the traditional account. Chisholm argued that in order
to understand the role of the agent, qua agent, in the production of an action we have to give up
the idea that only natural phenomena, such as desires and beliefs, can be a cause. This led
Chisholm to formulate the idea of ‘agent causation’: the idea that the agent is the cause of his
own behavior in a way that is not reducible to any natural event. Velleman, to his credit, is not
prepared to cross the Rubicon. He maintains that we should look for an explanation of agency
which gives the agent his proper place and is compatible with the scientific image of the world
(p. 130). For Velleman this means that we have to find something like a desire that fulfills the
functional roles that we identify as those of the agent: we need to find the desire that is the agent
so to speak (p. 137).

This is a far-reaching conclusion and one might think that one need not go so far as to
reduce the agent himself to a desire or set of desires. Perhaps it is sufficient to reduce only the
causal influence of the agent to some desire(s). For example, Harry Frankfurt argued that we can
understand this influence in terms of the agent’s second-order desires, desires that have other
desires as their object. We desire that some of our first-order desires are effective and will
determine our actions. Other first-order desires are not endorsed in this way. Frankfurt believed
that this process of endorsement is what constitutes the agent. When we endorse a particular
desire, we identify ourselves with that desire. This act of identification is where the agent comes
into the story.

A standard objection to Frankfurt’s theory is that it does not distinguish sufficiently among
the second-order desires. On Frankfurt’s view, any endorsement of first-order desires by second-
order desires counts as agency. However, just as the agent can be alienated from his first-order desires, he can be alienated from his second-order desires (p. 134). In the example of the slip of the tongue, it could be the case that my father is utterly bored with life or that he suffers from a major depression. Consequently, he could come to regard the unconscious arbitrariness with which he addresses his daughters as fitting with his assessment of the human condition. My father’s mixing up of the names of my siblings satisfies the conditions of Frankfurt’s theory. It is caused by a desire and he endorses this desire. However, in this scenario my father does not act. If anything, it expresses a lack of will on my father’s part. His mixing up of his daughters’ names (and his endorsement of this) is a sign of ennui or depression, rather than an action (p. 13). Note that this criticism is possible precisely because the agent is not identical to a desire or set of desires on Frankfurt’s account. This explains why Velleman aims to include the agent himself in the reduction of agency.

Velleman thinks Frankfurt’s theory is an improvement on the traditional account in that it requires the agent to be aware of his behavior. Thus, whatever the correct reduction of agent causation is, it will have to constitute some form of self-awareness. Velleman thinks that in addition to such self-awareness, the reduct will have to fulfill the functional role of the agent. It will have to determine behavior, to control and scrutinize one’s actions. What kind of desire could that be: a desire that amounts to a form of self-awareness and self-determination? Velleman’s answer is that it is the desire for self-awareness: the desire to know what one is doing. Fulfillment of this desire obviously will count as self-awareness. It also leads to self-determination. Agents who know what they are doing will do those things that they have decided to do. A decision to $\phi$ comes with the expectation that one will $\phi$. If one subsequently would not $\phi$, one clearly does not know what one is doing (p. 26). (One surprising conclusion that Velleman draws from this model of self-determination is that decisions to act are essentially self-fulfilling predictions (chapter 2). I will not discuss this here.)
The objection against Frankfurt’s model – that the agent can fail to endorse her second-order desires – indicates that the desire to know what one is doing cannot itself be something that can come up for scrutiny by the agent. It should be a pre-condition for agency. Otherwise, the agent might fail to have this desire without any impediment on her agency. Therefore, the desire for self-awareness has to have a special status among the agent’s desires. Velleman makes much of the status of this desire. According to Velleman, the desire to know what one is doing is a constitutive aim. A constitutive aim is an aim that characterizes behavior in so far as it is to count as a specific kind of behavior. For example, suppose I am engaged in a competitive game. My behavior is competitive, if and only if it is my aim to win. Once I no longer aim at winning, my behavior is no longer competitive. Velleman thinks that action in the full-blooded sense is characterized by the constitutive aim of knowing what one is doing. A behavior is an action if self-awareness is its aim.

I suspect that many are unaware that we have a desire to know what we are doing when we act. We act on considerations that move us towards certain outcomes. The fact that we are aware and in control of what we are doing – at least in the ideal case – is not the satisfaction of a desire we actually have. Rather it seems a by-product of the exercise of the capacity for action. What makes Velleman think we actually have this desire? His answer is the following. It need not be the case that the agent has this aim under that description. In the example of the role of the aim of winning in a competitive game, I need not be moved by the de dicto desire to win. I could have this aim if I am moved to perform better than my opponent does in this game. That aim amounts to having the aim to win – though not under that description. Applying this to action, Velleman says that in all instances of full-blooded action we need not have the de dicto desire to know what we are doing. Rather, in the case of full-blooded action, the considerations that move us amount – among other things – to the aim of knowing what we are doing.

I have a slight worry about this argument. If this is how we should think of the aim of self-awareness, we should accept that we have many additional aims. Suppose that I am thirsty and
realize that drinking a glass of water will quench my thirst. Next, I move into the kitchen, fill a
glass from the faucet and drink its contents. This behavior is an instance of action. The
realization that I am thirsty, the realization that I need a glass of water and my subsequent
moving to the kitchen are all such that I am fully aware of what I am doing. Thus, these
considerations constitute self-awareness. However, that is not the only thing they constitute. The
same considerations constitute a modicum of instrumental rationality. Does this mean that I also
have the aim of acting rationally to some degree? Similarly, these considerations constitute a
concern for my physical wellbeing. Do I therefore have the aim of physical well-being? In short,
just because we can re-describe certain motives in terms of other motives, it does not mean that
we actually have all these motives. Velleman’s move leads us into assuming a true plethora of
motives and aims, none of which are introspectively ‘there’ when we act.

In the essays, Velleman gives different formulations on the constitutive aim of action. In
The Possibility of Practical Reason (chapter 8), he argues that the constitutive aim of action is
‘autonomy’; in What Happens When Someone Acts? (chapter 6) it is the aim of ‘acting in accordance
with reasons’; in The Story of Rational Action (chapter 7) it is the aim of ‘acting on preference sets
that make sense’, while in the Introduction it is the above-mentioned aim of ‘knowing what one is
doing’. In the introduction, Velleman indicates that his thinking has developed over the years
since Practical Reflection. It is tempting to think that these are all different formulations of the same
motive. However, it is unclear whether they really are. Velleman could have been clearer on this
issue in his otherwise crystal-clear introduction.

Velleman’s way of thinking about agency – that the conditions of agency are met in so far
as one has the constitutive aim of action – has several implications. One surprising implication
has to do with the status of reasons for action. Bernard Williams has argued that (1) a reason is a
consideration that could move the agent to action. (2) Since only (actual) desires can move an
agent to act, it follows that (3) the (actual) desires of an agent are his reasons. (Williams’
argument is subtler than this, but the details are not important here.) Thus Williams is skeptical
about the idea that the agent can have ‘external reasons’, reasons for doing something
independent of what she actually desires. In Williams’ terminology, only ‘internal reasons’ can be
real reasons. One implication of Williams’ view is a complete reductionism of reasons for action
to motivation: to claim that someone has a reason to $\phi$ just means that this person desires $\phi$.

Velleman accepts Williams’ premises but rejects this reductionism. Not all desires cause the
agent to act in the full-blooded sense. Only desires that amount to having the constitutive aim of
action, or are compatible with the fulfillment of that aim, will count as reasons. The constitutive
aim of action, therefore, functions as a constraint on (1). Thus, Velleman can be an internalist,
without thereby being committed to reductionism about reasons. Velleman defines reasons for
action as follows:

The considerations that qualify as reasons for doing something are considerations in
light of which, in doing it, the subject would know what he was doing. They are, more
colloquially, considerations in light of which the action would make sense to the agent.
(p. 26).

Recently, Philip Clark (“Velleman’s Autonomism”, *Ethics* 111, no. 3: 580-593) objected to
this argument. Clark responded to the way this argument was formulated in *The Possibility of
Practical Reason* (chapter 8). There, Velleman formulates the argument slightly differently. First, he
identifies autonomy – understood as the conscious control of one’s behavior – as the
constitutive aim of action. Secondly, rather than saying that behavior counts as action if the
motives for the action constitute having the aim of autonomy, he claims that behavior is action to
the extent that it successfully attains this aim. One of Clark’s objections is that this implies that
every autonomous action is rational. Let me explain. Velleman accepts premise (1): a reason for
action is a consideration that moves the agent to action. Velleman is also a naturalist and
therefore believes that only desires can cause behavior (premise (2)). Behavior becomes action
only if it is successful in terms of its constitutive aim (here conscious control). This gives
Velleman the room to argue that not all desires count as reasons. However, it also means that
those desires which constitute the aim of autonomy and successfully cause the actions are in fact
the agent’s reasons. Thus, whenever an agent acts, he acts rationally – that is, on the reasons that he has. While this not complete reductionism about reasons for action, it comes very close. There is no room for saying that someone acted unreasonably or irrationally. If the agent acted he thereby acted on the reasons there were.

In the introduction, after mentioning Clark’s critique, Velleman explicitly states that he no longer endorses the view that autonomy is the constitutive goal of action. Rather it is the aim of knowing what one is doing. Velleman does not show how this would meet Clark’s criticism.

Here is how I think the argument should go. The role of self-awareness is analogous to the role of truth in belief as Velleman describes it (especially, chapter 10). Belief aims at the truth in two ways. First, belief is an attitude that accepts the state of affairs it represents as a fact. (This is Velleman’s preferred way of saying that beliefs have a world-to-mind fit). Secondly, it is constitutive for an attitude with that direction of fit to count as a belief if the acceptance of facts is regulated in ways designed to promote the acceptance of truth (p. 113, 250-252). This is what distinguishes believing that \( \phi \) from, say, assuming or fantasizing that \( \phi \). Beliefs can be real beliefs – provided they have the proper kind of acceptance – yet fail to be true. Therefore, in the case of belief we have a standard – truth – that functions both as the constitutive aim of belief and as a standard for its success as a belief.

The suggestion is that the constitutive aim of self-awareness plays a similar role in the case action. For example, we could imagine that an agent was moved by certain considerations in favor of \( \phi \) with the aim of knowing what he is doing (i.e., \( \phi \)). Suppose, however, that there are also considerations in favor of \( \psi \), which, if he acted on them, would make him more aware of what he is doing. This makes the aim of self-awareness both a constitutive aim and a standard for its success. Behavior counts as action if it has the aim of self-awareness. An action is reasonable or rational in so far as the agent is successful in attaining that aim.

Thus, it seems that there is room for criticizing the agent’s actions. Not every action is ipso facto rational. If this is indeed the way Velleman wants to argue, he can avoid skepticism about
reasons for action. Self-awareness on this account is a substantive standard which functions as a standard of rationality, independently of the desires of the agent in the assessment of the actions of the agent. I doubt that anyone in the grip of a less intellectual view of reasons for action will be attracted to this view. Self-awareness is an extremely minimal standard. Saddam Hussain knew exactly what he was doing when he ordered his elite troops to use chemical weapons on the Kurds. I find it hard to accept that this is sufficient to make Saddam’s actions rational. Velleman, on the other hand, seems quite happy with this conclusion. After all, he explicitly rejects any connection between the good and reasons for action or between the good and agency (chapter 5).

Velleman’s main contribution in this volume to the development of a plausible theory of action is that he gives us in straightforward naturalistic terms a way to make sense of the Kantian idea that practical rationality is a constitutive feature of agency. There is no need to refer to question-begging conceptions of autonomy or metaphysically ‘odd’ facts about our own nature to see that agency requires (some amount of) rationality. What is disappointing, however, is that there seems to be little room, if any, for more substantial, let alone moral, restrictions on what counts as good reasons.

Velleman writes crisply and clearly but it is never easy to follow him. It is only after reading most of these essays that one gets a grip on the overall picture and the concerns that motivated Velleman. The views in these essays are original, challenging and extremely well thought out. In addition to the problems that I discussed above, there are beautiful contributions on the problem of free will, the notion of well-being, the theory of collective action, the psychology of desire and belief as well as the status of formal theories of rationality. I recommend this book to anyone interested in action theory, philosophy of mind, meta-ethics, moral psychology and practical rationality. It is very much worth the effort (for that it is!) of studying.

Bruno Verbeek