Reasons and Action Explanation
Benjamin Wald and Sergio Tenenbaum

Introduction

In providing an intentional explanation of action, we cite the reasons for actions. Since Davidson’s seminal “Action, Reasons, and Causes” (Davidson 1967) the relation between these reasons for action and the explanation of intentional action has been at the forefront of philosophy of action. Davidson’s answer to the question was, at least in broad outlines, widely embraced and it very quickly became the “mainstream” view in action theory. In a nutshell, according to Davidson, a reason for action both rationalizes and causes the action; this view has become known as the “standard account” or the “standard story” of action. Davidson immediately realized that such an account cannot easily become a reductive account of intentional action. Actions can be caused by beliefs and desires that rationalize the action while obviously failing to be a case of an intentional action done for that reason; in these cases the reasons that potentially rationalize action are said to cause the action in a “deviant” way. The problem of deviant causation has been one of the more serious obstacles for a proper causal theory of action. In this paper, we first examine the advantages and difficulties of the causal theory. Investigating the problem of deviant causal chains reveals that there are in fact two related problems faced by a causal theory of action. First, the problem of deviant causation shows that it is difficult to come up with a reductive account of intentional action that understands intentional actions as an action that is caused by mental states. Second there is the problem that such a theory seems to fail account for how the agent herself is guided by her reasons, what we call “the problem of guidance”. Solving the problem of guidance will not on its own resolve the problem of causal deviance.

Section one: Causal Theories and the Problem of Deviance

One point of agreement among many otherwise disparate theories of action explanation is that action is explained by citing the agent’s reason for acting, offering what is referred to as a “rationalizing explanation”. However, it is less clear how the reason for acting serves to explain the action. What form of explanation is at work here? On the standard story introduced by Davidson, action explanation is a form of causal explanation, and thus that the agent’s reason for action plays a role in the causation of the action. On Davidson’s view, actions are explained by citing a belief
and a desire, or more generally a pro-attitude, that jointly cause and rationalize the action. We explain my going to the fridge by my desire to drink a beer, and my belief that there is beer in the fridge, which jointly caused by going to the fridge.

The causal theory of action has a number of benefits. It brings action explanation under the familiar heading of causal explanations, making clear how the agent’s reasons are relevant to the explanation of action. Furthermore, it makes clear how we can differentiate the various reasons that an agent might have had from the one he or she acted on. Suppose Jay, a Scrabble aficionado, needs to vote on which of the short listed candidates will be offered a job in the Philosophy Department at University U. Jane is clearly the best philosopher among the candidates, but she is also an excellent Scrabble player. Both the fact that Jane is an excellent Scrabble player and the fact that she is the best philosopher in the pool rationalize the action. If any consideration that rationalized (or justified) an action counted as a reason the agent acted on, we would have to say that Jay voted for Jane both because she was an excellent Scrabble player and because she was the best candidate for the University U position. But it is at least possible that only one of them was his reason to vote for Jane; that is, it is possible that only one of them explains his voting for Jane. For instance, Jay might have taken no account of Jane’s Scrabble skills, or he might have been so keen on having a suitable Scrabble opponent that no other considerations made a difference to his vote. The causal theory of action has an easy answer to the question “Which of the considerations that rationalize my action were the reasons that I acted on”. According to the causal theory, only belief-desire pairs that actually caused my action explain what I did. If we can have more than one reason for action, causal accounts seem well positioned to explain both how this is possible and the different extents to which the different reasons explain our actions (Cf. Arpaly and Schroeder 2014).

Some philosophers have accused the causal theory of action of psychologism (Cf. Dancy, 2003). That is, the causal theory seems to imply that only my mental states can be reasons for action. But psychologism seems obviously false: the fact that, for instance, my best friend is in London can be reason for me to go to London, but my friend being in London is a fact about my friend, not about my mental states. However, it is a mistake to think that the causal theory of action requires that all our reasons for actions be mental states. A causal theory of action requires that the existence of certain mental states is implied by each genuine case of intentional action, but it is not committed to the view that reasons are mental states (Cf. Davis 2010, Setiya 2007). Indeed, Davidson himself never says that the reasons for which we act are mental states. He says that primary reasons are mental states, and that “In order to understand how a reason of any kind rationalizes an action it is necessary and sufficient that we see, at least in essential outline, how to construct a primary reason” (Davidson, 1980, pg.4). “Primary reason”, however, is a technical philosophical term, and need not line up with our everyday notion of the agent’s reason. In other words, the reasons for which we act need not be mental states, but

---

1 If only events can be cited in causal explanations, then we can reframe this explanation in terms of the event of coming to have the relevant belief and desire.
they will rationalize, and hence explain, our actions via primary reasons, which are belief/desire pairs. My reason for going to the fridge can be to get a beer, or perhaps that there is beer in the fridge, and neither of these name mental states. Still, my reason only rationalizes my action because it is represented by my belief and desire, and these cause my actions. Mental states are not themselves our reasons for action, but they are what allow facts about the external world to be our reasons for action.

However, the causal theory of action faces a significant challenge; namely, the possibility of deviant causal chains. A deviant causal chain occurs when a belief and a desire cause a bodily motion, and the bodily motion is of the right sort to be rationalized by the belief and desire, but the way in which the causation occurs seems incompatible with the agent having acted, or at least with the belief-desire pair in question being the primary reason for the action. Davidson’s example is of a climber who desires to relieve himself of the burden of a fellow climber who he is supporting via a rope, and believes that he can relieve himself of this burden by loosening his grip on the rope. His coming to have this belief and desire makes him so nervous that he does end up loosening his grip on the rope (Davidson 1980). Here the belief and desire that would rationalize the action have indeed caused his bodily motion, but they have done so in a non-standard way. Intuitively, the climber’s beliefs and desires in this case do not provide a rationalizing explanation of the action of letting go of the rope; in fact, it seems that ‘letting go of the rope’ should not be classified as an intentional action at all. The challenge is to develop a causal account that does not incorrectly classify the bodily motions that stem from deviant causal chains as actions. Davidson himself did not think that the problem of deviant causation could be solved but he did not think that this was a fatal blow to the theory. According to Davidson, the belief and desire must cause the action in the right way, but he claims that he “despair[ed] of spelling out ... the way in which attitudes must cause actions if they are to rationalize the action.” (Davidson 1980: 79) However, unless we can specify of what counts as “causing in the right way”, we cannot avoid the suspicion that these theories are leaving out a central part of an account of acting for reasons: how reasons guide the agent in intentional action.

The literature on deviant causal chains is huge, far too large to be discussed in detail here. However, let us briefly consider some of the proposed solutions. Christopher Peacocke has provided an attempt to give a purely causal specification of the right kind of cause. On his view the intention must not just cause but also differentially explain the action, a suggestion also adopted with further caveats by John Bishop (Peacocke 1979, Bishop 1989). A cause differentially explains its effect when, roughly, there is a function that relates a feature of the cause to a feature of the effect. For instance, there is a function that maps the degree to which a dimmer is rotated to the amount of current that travels through a lightbulb, and hence to the amount of light produced, and so the degree to which the dimmer is rotated differentially explains the brightness of the bulb. More formally, we have a simple case of differential explanation when the following formula holds:

\[ (\forall x)(\forall n)(\forall t)((F_{xt} \& G_{xnt}) \supset H_{xk}(n)(t + \delta t)) \]
where \( n \) ranges over numbers, \( t \) over times, \( k() \) is a numerical function. Consider an ordinary action- I take a drink of water. Then \( x \) is a drinking of water \((F)\) and \( x \) is caused by a desire to drink water \((G)\) with an intensity \( n \) at time \( t \). \( x \) is then also a drinking of a specific quantity of water \((H)\) at time \( t + \delta t \), where the quantity drunk is determined by a mathematical function on the intensity of the desire \((K(n))\). Thus, since the desire differentially explains how much water I drink, it causes my action non-deviantly. On the other hand, consider a case of deviant causation such as Davidson’s climber. According to Peacocke and Bishop, we would explain the deviancy of this case by the fact that the climber’s desire to rid himself of the excess weight does not differentially explain any feature of his dropping of the rope. He drops the rope through nervousness, and it seems reasonable to hold that the precise way in which he drops the rope, the speed with which he opens his hand and so on, is not sensitive to any feature of his desire; it would have been identical even if it had been the result of a slightly different desire.\(^2\)

However, it is that it is not clear that this account succeeds in ruling out all cases of deviant causation. All that differential explanation requires is that there be some function relating features of the cause to features of the effect. But there may well be some such function even in cases of deviant causation. Perhaps the strength of the desire, which we could express numerically, determines via a numerical function the intensity of the tremors that shake the climber. Indeed, as Scott Sehon (1997, 207) points out, the function could just be a constant function, mapping every variety of nervousness to the same dropping of the rope. The existence of such a numerical function would clearly not render his action intentional.\(^3\)

In response to worries like this Peacocke adds the requirement to state that the cause must strongly differentially explain the effect, where strongly differential explanations are such that we can recover the conditions of the cause from the effect; in other words, that the function relating cause to effect be one-to-one (Peacocke 1979, pg.79-80). However, it is not clear that this helps. It is certainly conceivable that in a particular case of deviant causation each possible desire is paired up with a slightly different bodily motion. Furthermore, this requirement seems too strong. It seems perfectly possible that slightly different desires should result in the same bodily movement with the outcome still being an intentional action. To take Sehon’s example, imagine a pitcher with great control over the speed of his pitches, and a maximum pitching speed of 75 mph. His intention for how fast to throw the ball will differentially explain the balls speed. If he intends to throw it 60 mph, the ball will move 60 mph, if he intends to throw it 40 mph, it will go 40 mph, and so on. Now imagine that he intends to pitch the ball at 75 mph, and does so. The problem is that, since his top speed is 75 mph, had he intended to pitch the ball at 80 mph, it still would have gone 75 mph. So the function from his intentions to the balls speed is not one-to-one, but it still seems clearly intentional (Sehon 1997, pg.209). The Peacocke/Bishop account seems to getting at something

---

\(^2\) Bishop adds some extra caveats to deal with cases where the causal chain leads through other agents; since our objection does not deal with such cases we will not consider these additions here.

\(^3\) This criticism is made in more detail in Sehon (1997)
important missing from Davidson’s view: it’s not clear how in Davidson’s account the agent is *guided* by her reasons (more on this below). The Peacocke/Bishop might seem to give a more robust account of some sort of guidance; differences in the content of the agent’s attitudes will result in systematic differences in the action. However, the objections above suggest that it cannot capture this idea properly.

Kieran Setiya offers a promising solution to causal deviance, invoking exactly the idea of guidance (Setiya 2007, pp. 32-3; see also Frankfurt 1978). According to Setiya, what distinguishes a deviant cause from a non-deviant cause is that a non-deviant cause not only initiates an action, it continues to guide that action throughout its performance. As Setiya puts it, “if I do something with the end of doing, I must have a plan for doing $\phi$, by performing that action, and I count as doing $\phi$, intentionally just in case I do it in accordance with my plan” (Setiya 2007, 32). So, for the climber example, the problem is that the momentary loosening of one’s grip from nervousness may cause one to let go of a rope, but it does not guide this bodily motion in terms of any plan.

We might worry about how to apply this account to some very short action; how exactly does one guide the letting go of a rope once one has let go? And couldn’t one change one’s mind almost immediately, and so never actually guide the action after it is initiated, and yet have intentionally acted for the brief period of time preceding the change of heart? But there is a more important worry here. Let us start by asking what is meant by guidance in this account. It must be something more than just a requirement that the disposition persist in causing behavior over time. After all, the trembling climber, whose mountain climbing looks increasingly ill advised, could well be so nervous that his hands tremble for a prolonged period of time, thus ensuring that he cannot grab the rope once it has begun to slip from his grasp. We can add a feedback mechanism to this example as well. Perhaps the realization that he is dropping the rope makes his desire to be rid of the weight more vivid, which in turn increases his nervousness and this is what causes the trembling to continue. This is still a case of deviant causation, so Setiya must hold that this is not sufficient for the climber to count as having his trembling be guided by the goal of ridding him of the weight of the other climber.

We need the notion of guidance to be more robust. Setiya does not offer an explicit account, saying “sustained causation of a process towards a goal is not unique to intentional action: it is present in purposive behaviour that is not intentional. So although it is something of which we lack an adequate theory, there is no circularity in taking it for granted here” (Setiya 2007, 32). It is true that there is no circularity, but even if we take for granted some understanding of sustained causation towards a goal, it is far from clear that we can use it to draw some kind of deviant/nondeviant distinction, or that it will track the relevant distinction for intentional action. When the heart pumps blood to various organs, for example, it not only initiates but also sustains this process. However, it is still possible for this

---

4 This account is intended to apply only to basic actions, with complex actions give a different treatment. We shall focus on the basic case here.
kind of sustained causation to admit of a difference between deviant and non-deviant forms of causation. A case in which one’s heart “malfunctions” but fortuitously still causes blood to be pumped, even if such “malfunction” continuously causes blood to be pumped, will be a case of deviant causation. It would certainly be a case of deviant causation, and not of the right kind of sustained causation, if, for instance, the cells in my heart stopped beating in unison and instead began to beat in a complicated musical harmony with one another, a musical harmony that alien observers found pleasant and which therefore caused them to use their advanced technology to circulate blood through my body in order to continue enjoying the beat. My heart would still be sustaining the circulation of blood through my body, but this would be a highly deviant causal chain; my heart would not be instantiating its function of pumping blood or be performing its characteristic activity. Thus, the deviant/non-deviant distinction here is not accounted for in terms of whether or not we have sustained causation towards a goal. Of course, one might deny that such fortuitous “malfunction” cases count as cases of sustained causation towards a goal, exactly because the heart is not pumping blood in the right way. But, absent a more principled account of what the “right way” consists in, this response would risk explaining sustained causation in terms of nondeviance rather than the other way around.

One might hope that a principled account of the right sort of causation will be provided by another area of philosophy. For example, to provide an evolutionary explanation of the development of the heart, we might need to appeal to sustained causation towards a goal, and differentiate deviant from non-deviant versions of this form of sustained causation. However, even if this is true, it isn’t clear that Setiya can help himself to just any such account of sustained causation in answering the issues arising in philosophy of action. It is plausible that the account of sustained causation in evolutionary biology will appeal to the evolutionary history of the organism. Perhaps the heart pumps blood non-deviantly when it pumps blood via causal mechanisms that are typical of the heart’s evolutionary history. The problem is that there could be causal processes involving desire-belief pairs that were selected for, and thus nondeviant from the point of view of natural teleology, but that were no less deviant from the point of view of intentional explanations. For example, perhaps an agent is so constituted that whenever they desire to avoid a predator and believe that they could do so by falling to the ground, all of their muscles are paralyzed and they fall to the ground. This might be a case of “sustained causation” such that as long as the belief is occurring, the effect is still on, and the agent’s muscles continue to be paralyzed, and it is evolutionarily selected for a goal, avoiding the attention of predators. This tendency could well have been evolutionarily selected for for precisely this purpose, but even if it were it would not render the causation any less deviant from the point of view of action.

In fairness to Setiya, his causal theory, unlike Davidson’s, does not take belief/desire pairs to be the mental state that cause intentional action, but rather intentions understood as belief-like states that also motivate. However, the problem does not seem to depend on the kind of mental state in question.
explanation. In sum, it is far from clear that we can solve the problem of deviant causation even if we help ourselves a notion of sustained causation available in the context of natural teleology; it is not clear how sustained causation can single out the cases in which the agent was being guided by her reasons to act, except by adding something like “the right kind of sustained causation”.

The difficulty of resolving deviant causal chain worries suggests that this is no minor technical issue, but reflects some deeper feature of the explanation of action. David Velleman argues that what is missing is from the standard causal account of action is that “In this story, reasons cause an intention, and an intention causes bodily movements, but nobody—that is, no person—does anything. Psychological and physiological events take place inside a person, but the person serves merely as the arena for these events: he takes no active part” (Velleman 1992, 461). In fact, this problem with causal accounts point to a tension for any account of action explanations that arises from the very intuitive idea that intentional actions are such that an agent is guided by her reasons to act. We can separate two independent problems that the possibility of deviant causal chains raise for standard causal theories of actions. First, it raises doubts about whether the causal theory of action succeed as a reductive account of intentional action; the need to appeal to “causation in the right way” suggests that the reduction failed. We will call this problem the “proper problem of deviance”, or “the problem of deviance” for short. Second, they raise doubts about whether, at least on its own, the standard causal theory can succeeded as an account of intentional action, or whether it leaves out an essential aspect of intentional agency; whether it can explain not only how agents are moved by their reasons but guided by it. Standard causal theories do a good job in capturing how the agent’s reasons are involved in her action, but then they have difficulty explaining the role of agent; in Velleman’s words, the agent seems to take “no active part” in the action. Standard agent-causal theories suffer from the opposite problem. Even if we ignore metaphysical qualms about agent-causation, theories in which the agents’ causal powers are not determined by the causal powers of mental states such as belief, desires, and intentions will have a problem explaining how the fact that (I believe that) x is a reason to Φ is relevant to the explanation of my behaviour.6 We will call this “the problem of guidance”.

In the next section, we will show more precisely how attempting to capture this form of agential involvement creates a dilemma for theories of action explanation, and that this dilemma is not restricted to causal theories, but is a problem for all theories of action explanation. We then propose a specification of how an agent must be present in an action for it to count as an action, and thus a suggestion as to what causal theories may be missing as well as what they are getting right.

Section two: Being Guided by Reasons

---

6 Chisholm (1964) wrestles with a similar question.
We have suggested that the possibility of deviant causal chains is due to the absence of the agent him or herself from the production of the action; causal theories take the agent to be *responding* to reasons, but can’t explain how the agent is *guided* by reasons. The action must not just be a response to the agent’s reason, the reason must guide the agent. As Korsgaard puts it:

Neither the joint causal efficacy of the belief and the desire, nor the existence of an appropriate conceptual connection between them, nor the bare conjunction of these two facts, enables us to judge that a person acts rationally. For the person to act rationally, she must be motivated by her own recognition of the appropriate conceptual connection between the belief and the desire. We may say that she herself must combine the belief and the desire in the right way (Korsgaard 2008, 63)

It is worth comparing this concern with the issues that philosophers inspired by Frankfurt’s seminal work (Frankfurt 1971) have raised. If the desire that causes the agent to act is not one that the agent identifies, the action will not be full-blooded, or autonomous, or fully intentional. We need to identify a motive that, when it is involved in the production of action, constitutes the agent’s active participation in the action. But what kind of motive could do this job? For any candidate motive, we might worry that the agent might find themselves alienated from this form of motivation, and hence this motive would not represent their agency in that instance. Velleman characterizes the issue, roughly, as trying to find a mental state such that when it acts, the agent acts. (Velleman 1992). The problems might seem quite different, but we can see that these are similar worries about the role of agent in a purely causal understanding of intentional action: Korsgaard is concerned that rational agency requires that the *agent herself* “must combine the belief and desire”, while Velleman is concerned that we find a motive whose presence in action guarantees that the agent him or herself is active in the production of the action.

One possible solution to the issue of agent involvement and guidance by reasons is to identify a constitutive aim of agency or form of agency. In a nutshell, we had difficulties understanding how merely responding to reasons could count as proper agency. But perhaps there is a way of responding to reasons that when we respond to a reason in this way, we are thereby acting. For instance, if seeking to maximize utility is constitutive of agency, then one can think that if one is responding to a reason because (one takes that) responding to this reason maximizes utility, one is thereby acting. The presence of this motive will ensure the agent’s active participation in the production of the action, since the motive will be constitutive of agency; the agent is not a passive observer of her actions for there is no conceptual distance between the motive producing the action and the action being brought about by the agent herself.7

---

7 Velleman uses this claim to support his constitutivist metaethical position, on which normative
Velleman suggests a candidate motive, that of making sense of our own actions, but we can also see other theories of agency as providing similar, though often nonreductive, accounts. For example, theories of agency that endorse the “guise of the good” thesis hold that for something to be an intentional action it must be motivated by something like an agent’s “take” on the good (Cf. Davidson 1980, Raz 2002, Tenenbaum 2007). Guise of the good theories thus hold that the motive of doing good is constitutive of agency.

Such theories also have additional resources to resolve the problem of deviance. Davidson’s climber lets go of the rope under the causal influence of a relevant belief and desire. However, perhaps the explanation of the fact that this fails to be an intentional action is that the further motive constitutive of agency is absent from the causal story. For example, the motive of doing what is good does not seem to have any causal impact on this behavior. If the climber judged it good to let go of the rope, and this judgment causally contributed to his letting go of the rope, then perhaps we have a case of action after all. Or perhaps what is missing is the causal influence of a motive to satisfy as many of one’s desires as can be mutually satisfied. Where such a desire to be causally efficacious, it would mean that the agent judged that dropping the rope would contribute to satisfying as many of one’s desires as possible, and perhaps this is sufficient to render this an action. Of course, the details of this strategy will depend on what motive or aim we take to be constitutive of agency, or what we take to constitute the form of agency. But accepting some such motive offers us a promising route to addressing the worry of deviant causal chains, by providing a role for the agent’s active participation in the production of the action.

This solution won’t, on its own resolve the problem of deviant causal chains. Even if we add a further motive that is constitutive of agency, there could be further intervening states between this motive and action. Let us take for instance the constitutive motive in Velleman’s early account; namely, the motive of self-understanding. There could, it seems, be a case in which the motive of self-understanding is causally relevant and yet the action is still generated in a circuitous manner. Still, this strategy might provide additional resources for a solution to the problem. Perhaps the counter-intuitiveness of counting some cases of deviant causal chains as actions could be shown to derive from suspicions about guidance, rather than causation per se. One might insist that as long as the motive causes and maintains the action (and satisfies whatever other conditions the theory imposes) then we have a genuine case of intentional action. But even if this strategy is not successful, the constitutive aim seems to address at the least the problem of reasons are constructed out of the constitutive aim of agency. However, accepting that agency is constituted by the presence of a certain motive does not commit one to a particular metaethical position. One can adopt Velleman’s suggestion about action explanation without becoming a constitutivist. On constitutivist views, see Paul Katsafanas, “Constitutivism about Practical Reason” in this volume.

8 Davidson (1980) is arguably defending a somewhat different view, but we leave these complications aside.
guidance. As Velleman argues, even among cases in which the belief and desire causes the action in ways that are otherwise nondeviant, we can still have cases in which we have something that falls short of ‘full-blooded’ intentional action, exactly because the agent is not in control of how her desires are issuing in action.\(^9\)

However, in addressing this problem, this approach brings with it new challenges. We must determine how the motive that is posited fits into the agent’s reason for acting. An advantage of the causal account of action, recall, was that it allowed us to determine the agent’s reason for acting from all of the possible motivations that might have been at work. On the current account, however, the key role is played by the motive constitutive of agency. After all, it is the contribution of this motive that transforms a bodily motion into an action. Thus, the agent’s reason for action seems to be given by this constitutive motive, and is the same for every action. Take as an example the view that the motive of doing good is constitutive of agency. Then it is the belief that the action is good combined with this motive that renders any bodily motion an action. Thus, it seems that our reason for acting is the same for every action - the goodness of the action.

This position seems to have a number of implausible consequences. An obvious problem with this view is just that we seem to act for more than one reason. What use would it be to ask someone why they did something if the answer would always be the same? (Cf. Korsgaard 2008). A related, but more serious, problem for this view is that it renders agents problematically self-absorbed. Consider the view that the constitutive motive is the motive to do what is good or what one has sufficient normative reason to do. In that case, the agent’s reason for action always include that the action would be good or that one has sufficient normative reason to do. The problem is that, on this account, it seems that an agent always thinks “one thought too many”, to use Bernard Williams terminology (Williams 1981). The account seems to imply that no one ever jumps into the pool to save their drowning child for the reason that “my child is drowning”. They must always take the further step cognitive and act on the reason that “saving my drowning child would be good”. Douglas Lavin terms this “the problem of narcissism” (Lavin, 2011).\(^10\) The problem is that on this account of agency we are narcissistically focused on the goodness of our own actions. We seem to care more about the goodness of our actions than we do about the wellbeing of our friends and family, the outcomes of our projects, or the suffering of those we aid. We seem to face a dilemma here. If we focus on the reasons for which the agent act, and take them to be the primary explanans of the action, we run against the problem of guidance; it is not clear how the agent is being guided by the reasons, or how the agent is involved in the action. If we focus on the features that can potentially explain the involvement of the agent, such as a constitutive motive, we run against the problem of narcissism.

Of course, the dilemma is not an artifact of choosing the motive of pursuing the good as the constitutive motive. The situation is no better if we consider other

\(^{10}\) See also Korsgaard (2008).
putative constitutive motives. On Velleman’s suggested motive in his earlier work, that of self-understanding, agents end up narcissistically obsessed with the state of their own self-understanding, which is surely no better than being fixated on the goodness of one’s actions. Even if the constitutive aim is to maximize utility or act on the strongest desire (or set of desires), we face a similar problem. On this view, our reason for acting on a given desire is not the content of the desire, but the fact that it allows us to satisfy our strongest desire; the father who jumps into the water to save their child does for the reason that doing so allows him to satisfy his strongest desire, rather than for the reason that his child is drowning.

Moreover, the challenge is not restricted to theories that postulate a constitutive aim or form of action, but to any theory that distances itself from the model of understanding agency as a direct response to reasons. As we said above, the main advantage of the standard causal account is in explaining how the action is sensitive to reasons. A similar problem seems to be faced by, for instance, George Wilson’s account of intentional action and action explanation (Wilson 1989). According to Wilson, an intentional explanation does not posit a causal relation between a desire and an action (or more generally between motivating reason and cause). Instead, locutions such as the following express an irreducible form of rational teleology:

\[ S \text{ is } \phi \text{-ing in order to } \psi . \]

Wilson claims that this is what explanations by reasons is, and the fact that this is supposed to be a form of rational teleology guarantees that ‘\( \psi \text{-ing} \)’ is the motive or the reason why the agent acted. But this account makes it difficult to understand the relation between the action and the putative normative reasons that would justify such an action. Given that there is no causal relation between the action and the agent’s motivational state, it is not clear how the action could be responsive to the “good-making” or “reason-making” features of the action. Although we can say that, for instance, Jane gave a ride to Louis in order to help someone in need, it is not clear how we can say that Jane is responding to the good making feature of the action.\(^\text{11}\)

But perhaps the constitutive motive strategy can avoid our dilemma. The advocate of this strategy could argue that ordinary motives and constitutive motives play different roles in action explanation. In other words, one could argue that while the constitutive motive is indeed present in all our actions and plays a crucial role in rationalizing action, it does not do so by being an extra reason for acting. We want to preserve what we ordinarily take to be our reasons for acting, while still finding way of explaining how the agent is guided by the reason, rather than just being the locus of the causal interaction between an external fact and movement of her bodies. There are three basic ways of doing this. First, one could argue that the constitutive aim of action is not “the” reason we act for, but it is an extra motive, added to the other motivating reasons. This is arguably Velleman’s strategy in his early work.

---

\(^{11}\) Of course, the teleological account could try to provide an explanation of reason responsiveness; our only point here is that such explanation is needed.
(Velleman 1992). The other is a strategy employed by a number of advocates of the ‘guise of the good’, which we shall call the attitude strategy: the constitutive aim of action is not part of the content of the relevant attitude but it is constitutive of what the attitude is. The idea is not that attitudes that figure in the explanation of action, such as intentions and desires, have ‘good’ as part of what is represented in the attitude. Rather it is that in having these attitudes, irrespective of their content, one aims at the good. This would give the constitutive aim of the action a quite different role than the reasons for action. This is essentially the strategy followed by one of us (Tenenbaum 2007, 2008, and 2012) and Karl Schafer (Schafer 2013). Finally, one can try to understand the nature of the attitudes in question such that they explain what is being guided by a reason while not committing oneself to agency having any constitutive aim, or that the agent is pursuing anything other than the object of her reason. Thus there is no competing aim or end to displace what we ordinarily take to be the reason for action. This is essentially the strategy implicit in Setiya’s work (Setiya 2007). We will briefly discuss the first and third option, and then try to explain the motivation for the middle option in the next section.

Let us begin with the first option, embodied in Velleman’s invocation of the desire for self-understanding as a necessary element in explaining action in his earlier work. Velleman does seem to provide a way in which the constitutive aim of action plays a role in the production of action while preserving at the same time a sense in which the agent is responsive to the actual reasons. Without the participation of the motive of self-understanding, our bodily motions would not count as actions. But without the other sundry motives that move us to act, the motive of self-understanding would have nothing with which to work. We understand ourselves, according to Velleman, by ensuring that the motivations that best fit our self-understanding are the ones that move us to action. Without other motivations to favor or suppress, the motive of self-understanding would be empty. Thus, on Velleman’s account, there is clearly a causal role for the ordinary desires that we take to establish our reasons for action, and hence for our responsiveness to our ordinary reasons for action.

However, it is not clear that Velleman can escape the charge of narcissism by merely adding a role for ordinary reasons for action. After all, even if we make room for both ordinary desires and the desire for self-understanding, both desires still seem to be playing the same type of causal role. If the desire to help the poor and the desire for self-understanding play the same role in the etiology of action, it seems that the right answer to the question “Why did you give Louis a ride” is “Because Louis’s needed help and because if I did anything else my self-understanding would be compromised”. The fact that the constitutive aim is now alongside the reason to help is unlikely to be much consolation; when the worry is that one’s description of agency adds “one thought too many”, adding yet more thoughts (or more roles to

12 Another version of the view would claim that in having the attitude we aim to be responding to normative reasons.
other thoughts) is unlikely to be the answer.

There are differences between the motive of self-understanding and our ordinary reasons for action, and a defender of a position like Velleman’s could argue that this motive is sufficiently distinct in its operation that it ought not to be considered a reason for which we act. For instance, since it is by the participation of this motive that other considerations become reasons for which an agent (full-bloodyedly) acts, and since it seems senseless to say the motive participates in itself, it would be inaccurate to call the constitutive motive itself a ‘reason’. However, the challenge for one who wishes to pursue this strategy is to show how this move is more than a verbal stipulation. If the motive of self-understanding plays the same etiological role in bringing about action as do our ordinary reasons for action, and it contributes in an important way to rationalizing the action so brought about, it seems to be fulfilling all of the functions of the agent’s reason for action. Thus, what we need is an explanation of why this further motive is not an instance of the problematic “extra thought”.

A possible alternative approach is to reject the idea that the agent’s responsiveness to their reasons is embodied in a motive or goal at all. Instead, it may merely be a requirement that a certain sort of internal structure be present. For example, Kieran Setiya in his account of what it is to act for a reason rejects what he calls “teleological accounts” of acting for a reason. On Setiya’s account of acting for a reason, we begin with the traditional belief desire pair. This, however, is not yet enough to make the resulting behavior a case of acting for a reason. We must add a further mental state, a hybrid state that combines a cognitive component with a motivational component. This mental state is reflexive- it has as its content “that one hereby doing $\phi$ because of the belief that $P$” (Setiya 2007, 46). Acting for a reason is a representation of one’s action as being caused both by itself and by the belief that $P$. The motivational component is similarly two-fold. On the one hand, the attitude itself contributes motivationally to the production of the action. On the other hand, it also enables the motivational contribution of the belief that $P$. As Setiya puts it, “it is only when I take the fine weather as my reason for walking that my belief about the weather causes me to walk; until I do so, its causal power is muted. So taking something as my reason for doing $\phi$ is desire-like not only in causing my action, but in causing it to be caused by the relevant belief” (Setiya, 2007, 43).

Let us leave aside our earlier concerns we had with whether Setiya’s view can avoid the problem of deviance and focus on whether it escapes both horns of our dilemma. At first, it does appear to succeed. The desire-belief attitude seems to show how the agent is guided by her reason and involved in the action without displacing the agent’s reason for action. It is not an alternative goal or feature of the action that makes it the agent’s own- instead, the missing feature is a reflexive self-awareness of the sources of one’s own motivation. However, it does so by introducing two “becauses” in the explanation of the action; the agent acts because of the desire-like belief but also because of the specific belief that $P$. But it is not clear how this is possible; how it is possible for the desire-like belief to both cause
my action and to cause it “to be caused by the relevant belief”?\textsuperscript{13} Although obviously there are cases in which something causes a causal relation to obtain (when, for an instance, I fix my computer so that typing “t” will now cause a “t” to appear in the screen), these typical cases seem very different from the relation being proposed here. The causal relation posited here is at least obscure. It seems that the dilemma is resolved by stipulating that, rather than explaining how, the two states make distinct contributions to the explanation of the action.

These are hardly conclusive arguments against views such as Velleman’s and Setiya’s. But we hope they at least give us some motivation to look into the views of second kind in more detail.

\textit{Section Three: Action Explanation, Guise of the Good, and Reasons}

The most clearly developed examples of the attitude view arise in the context of theories of action that accept the guise of the good these, or “scholastic” views. According to such theories, action requires seeing the object of the action or desire as something in some way good. There are two kinds of scholastic theories (Tenenbaum 2007). On the first kind of view, ‘good’ is part of the content of the practical attitude in question (Oddie, 2005),\textsuperscript{14} or of another attitude, such as a belief, that is a necessary condition for having a certain practical attitude or for intentional action (Raz 2010). For instance, it may be part of the attitude of intention that the action intended is in good in some way. On the second kind of view, ‘good’ is not, at least not necessarily, part of the content of the attitude (Tenenbaum 2007, 2008, and 2012; Schafer 2013). According to the second view, it is part of the nature of the practical attitude that in \textit{having} the attitude one somehow judges the object of the attitude to be good, or is presented with its object as good, but that this judgment is not a part of the content of the attitude. We will be discussing only the second kind of view as it has a decided advantage for our purposes.\textsuperscript{15}

The upshot of part two, recall, was that we wanted to identify a constitutive aim or form of agency, in order to account for the agent’s active participation in the production of action. The problem is that adding such a constitutive aim seemed to put us on the second horn of our dilemma; if our view implies that the agent’s reason for acting is always given by the constitutive aim of action, we face the problem of narcissism. However, if the constitutive aim of the attitude is not part of the content, then the agent could be guided by the aim without representing it. Instead, just in virtue of having formed the requisite attitude, the agent will already

\textsuperscript{13} See (Pendlebury 2013) for a similar (and more detailed) criticism.
\textsuperscript{14} This is Oddie’s view for the case of desire.
\textsuperscript{15} The attitude strategy is not restricted to scholastic views. As we go on to show, it is plausible that the attitude of belief is characterized by a judgment that the content of the belief is true, and alternate views could hold that the judgment embodied in desire is not that the object of the desire is good, but some other judgment, perhaps that pursuing the object of the desire is conducive to self understanding, to adapt Velleman’s view. In what follows, we will focus on scholastic views for the sake of clarity. However, for all we say here the judgment embodied in the attitude view could be something quite different.
have been implicitly guided by the aim in question. On this proposal, when I act on
the reason that someone needs help, I don’t act also on the reason that I must
conform my action to the normative reasons, or the reason that I must do what is
good. In intentionally \( \phi \)-ing, I thereby take \( \phi \)-ing to be good or something that I
have normative reason to do (for the reason I \( \phi \)-ed). This is not a further judgment
or attitude alongside my intentionally \( \phi \)-ing, it is just part of what it is to \( \phi \)
intentionally. But this proposal also avoids the problem of guidance. In \( \phi \)-ing for a
certain reason I am not merely passively reacting to a certain fact or content of a
desire but I am also active: I am \textit{taking} \( \phi \)-ing to be good.\(^{16}\)

But how exactly taking something to be good (or taking it to be a normative
reason for action) be part of the nature of an attitude? We cannot give a full answer
to this question here, but we hope at least to show that there is very good reason to
think that this is a feasible option that should be taken seriously.

Let us start by taking a case that seems to be analogous (and that defenders
of the guise of the good view have explicitly put forward as a fruitful analogy; see
Tenenbaum 2007 and 2012 and Schafer 2013): the relation between belief and
truth. There is no question that ‘true’ is not part of the content of every belief; it is
not the case that if I believe that \( p \), the real content of my belief is ‘\( p \) is true’. This
would be a picture of belief that is every bit as unintuitive as the narcissistic account
of our reasons for action was. However, there seems to be an important connection
between an agent believing that \( p \) and that agent accepting the truth of \( p \); the
connection is generally referred to in the claim that “belief aims at the truth”. We
can think that the attitude view as claiming that a similar relation holds between
intention (or action and good).

It is a matter of great controversy of what exactly is the nature of the relation
between belief and truth (see, for instance, Velleman 1992, Wedgwood 2002, Shah
and Velleman 2005, Staglich-Peterson 2006). For our purposes, a couple of points
are particularly important. First, it is plausible to suppose that the relationship
between belief and truth is not merely normative; that is, the relation is not
exhausted by the fact that, in some sense, truth is a norm or ideal for belief.\(^{17}\) Let us
look at a variation of Moore’s paradoxical sentence: “‘It is raining’ is true, but I do
not believe it.”\(^{18}\) The person who says such a thing is not just being irrational.
There is nothing particularly puzzling about admitting to irrationality, even ongoing
irrationality; but it is not clear how this Moore paradoxical sentence can even be
asserted (For more on this point, see Tenenbaum 2012). The explanation for the
paradoxicality of this kind of statement seems to imply that believing is not just

\(^{16}\) Of course this would not (or least would not \textit{obviously}) generate reductionist account; there is no
guarantee that we can provide any kind of reductive analysis of the attitude in question

\(^{17}\) One of us presents a further argument that the relation between belief and truth is not purely

\(^{18}\) There are certain cases that this might be less paradoxical. If, for instance, the sentence to which ‘is
true’ is predicated is not in a language that the speaker can understand it, but that she knows to be true
through testimony and yet she knows that \( p’ \) expresses a proposition she does not know, then she
would be a position to assert such a sentence. We leave these complications aside.
being in a state which is “correct” if it is true, but instead believing just amounts, in some sense, to holding the content of one’s belief to be true. On this view, in saying “it is raining’ is true”, the agent expresses that they hold ‘it is raining’ true, while in denying that they believe that it is raining, they deny that they hold ‘it is raining’ true- thus accepting the sentence is akin to accepting a manifest contradiction. John Gibbons (2013) makes a similar point about belief using the notion of commitment. Believing that P commits us to the truth of P. Thus, discovering that P is false will also mean discovering our belief that P is defective from our own point of view.\textsuperscript{19} The idea of commitment is a helpful one, since it emphasizes the way that beliefs that fail to meet this standard are defective according to the agent’s own standards. However, normally I can be fully aware that I am failing to live up to my commitments; out of laziness, I might stay at home at Elections days even if I am committed to the importance of the democratic process and voting in particular. But I cannot, at least in the paradigmatic cases, be at the same time fully aware that I believe that p and that p is not true and just “feel guilty” about my failure live up to my commitment.

Although belief involves a taking of, or commitment to, P being true, this is not an aim or goal of belief in the ordinary sense; we do not form particular beliefs in order to have true beliefs. Similarly, it is not true that all reasons for belief that p are of the form “accepting p will bring it about that I have a true belief”; instead in accepting that q is reason to believe p, one thereby accepts that one holds p to be true on the basis of q. Thus, if reasons for action work in a similar way, then our reasons for action are not necessarily of the form “Doing φ will bring it about that I perform a good action”; instead, in accepting that P is a reason to φ one thereby takes φ-ing to be good on the basis of P.

Of course, even if we are right that this analogy with belief holds, we cannot consider that the job of the attitude view is done at this point. As we said, the relation between belief and truth is not well-understood; although it is comforting to know that other areas face similar tasks to the one faced by the attitude view, noticing this comforting thought cannot replace actually engaging in the task. The full vindication of the attitude view thus depends on whether it can provide a satisfactory account of the nature of an attitude such that in having them, one judges something to be good (or is presented with something as good). Here we can only outline some possible ways in one can provide such an account. The most obvious one is to think that given the roles of these attitudes in intentional explanations, they must be seen as necessarily expressing normative or evaluative stances. We may accept, for instance, Anscombe’s view that intentional explanations are answers to a particular question “Why?” (Anscombe 1957). In Anscombe’s view, not all possible contents of motivational attitudes are candidates for a proper answer to

\textsuperscript{19} Actually, Gibbons argues that the commitment of belief is stronger- in believing P we are committed to knowing that P. This will of course includes a commitment to P’s being true, but it also entails that learning that our grounds for believing P are defective will likewise forces us to revise our belief, just as learning that it is false will. In any case, we only need the link between belief and truth, and so we will leave aside the question of whether belief also entails that you take yourself to know that P.
this “Why” question. Only some of them would make either the action or the agent intelligible. If the conditions under which an answer to the Anscombe’s question is intelligible turn out to be that the answer allows us to understand the agent as pursuing her evaluative or normative attitudes, then we would have an argument for the conclusion that the motivational attitudes cited in the explanation express the agent’s normative or evaluative views.

A similar argument can be made if we accept Davidson’s view that intentional explanations (and, in Davidson’s case the mental more generally) are informed by the “constitutive ideal of rationality”. On one interpretation of this view, intentional explanations place the agent’s action within “the pattern of a life led by an agent who can shape her action and thought in the light of an ideal of rationality” (McDowell 1998, p. 35). If understanding an action as fitting into this pattern depends on seeing the agent as trying to pursue the good or act for good reasons, we would have the makings of a similar argument. A quite different strategy is pursued by (Boyle and Lavin 2010). Very roughly, they defend an Aristotelian conception of animal action as pursuing the good of the animal, and argue for the claim that intentional action is an essentially self-conscious case of animal action; that is, one in which one is guided by one’s self-conscious understanding of the human good. On their view, “the intentional actions of a rational agent express his regarding those actions as good in a plain enough sense: for such an agent has the power to reflect on how to act, and if on reflection he does not accept that a given way of acting has at least something good about it, he will in so doing have changed his mind about whether to do it” (Boyle and Lavin 2010, pg. 191).

We cannot adequately pursue any of these avenues here. But the fact that this position promises to solve the problem of guidance, and that a similar position seems to be also need in the realm of theoretical reason gives us good reason to think that it’s worth pursuing them.
Work Cited


Wilson, George (1989), *The Intentionality of Human Action* (Stanford University Press).