

# Acting for a Reason and Following a Principle

Andrew James McAninch

Accepted: 17 November 2014 / Published online: 26 November 2014  
© Springer Science+Business Media Dordrecht 2014

**Abstract** According to an influential view of practical reason and rational agency, a person acts for a reason only if she recognizes some consideration to be a reason, where this recognition motivates her to act. I call this requirement the *guidance condition* on acting for a reason. Despite its intuitive appeal, the guidance condition appears to generate a vicious regress. At least one proponent of the guidance condition, Christine M. Korsgaard, is sensitive to this regress worry, and her appeal in recent work to the constitutive principles of action can be seen, in part, as a response to it. I argue, however, that if we are to appeal to the constitutive principles of action to resolve the regress, then we must determine whether acting on such principles is also subject to the guidance condition. This raises a dilemma. If following these principles is subject to the guidance condition, then the regress remains unresolved. But if not, then the rationale for applying it to acting for a reason vanishes as well. I conclude that we should embrace an account of acting for a reason that rejects the guidance condition.

**Keywords** Agency · Constitutivism · Korsgaard, Christine M · Practical reason · Regress

## 1 Introduction

On an influential view of practical reason and rational agency, a person acts for a reason only if she recognizes some consideration to be a reason, where this recognition motivates her to act (see, e.g., Korsgaard 1997). Call this requirement the *guidance condition* on acting for a reason. Such a view is appealing in many ways, especially in its attempt to link reason-responsiveness to a distinctive kind of self-determination. The guidance condition is meant to secure an agent's active involvement in her own action, without which the action would not appropriately be attributable to her. Furthermore, were the agent's action not appropriately attributable to her, then we could not appropriately hold her answerable or responsible for her action. Some have raised the worry,

---

A. J. McAninch (✉)  
Department of Philosophy, University of Pennsylvania, 433 Cohen Hall, Philadelphia, PA 19104-6304, USA  
e-mail: mcaninch@sas.upenn.edu

A. J. McAninch  
andrew.mcaninch@gmail.com

Andrew James McAninch  
e-mail: andrew.mcaninch@gmail.com

however, that the guidance condition, with its requirement of reflective recognition, generates a vicious regress of reflective recognitional acts (e.g., Arpaly and Schroeder 2012; Railton 2004). At least one philosopher who endorses the guidance condition, Christine M. Korsgaard, appears to be sensitive to this worry. Indeed, her appeal in recent work to the constitutive principles of action can be seen, in part, as a response to this worry, as a way to halt the regress.<sup>1</sup>

I will argue, however, that this appeal to the constitutive principles of action fails to resolve the initial problem. If we are to appeal to the constitutive principles of action to resolve the regress, then we need an explanation of what following such principles requires. Specifically, we must determine whether the guidance condition applies to *following a principle* just as it applies to *acting for a reason*. This question raises a dilemma: if the guidance condition does apply to following a principle, then the regress remains unresolved, but if it does not apply to following a principle, then the motivation for applying it to acting for a reason vanishes as well. The upshot of the dilemma is that we should embrace an account of practical reason and rational agency that rejects the guidance condition. But doing so, I will argue, does not imply that we cannot be answerable for our actions in the way that matters in our moral and epistemic practices.

I will begin, in the next section, by describing the initial regress problem that the guidance condition on acting for a reason raises. In the third section, I will explain how Korsgaard's appeal to the constitutive principles of action is meant to resolve the initial regress problem. And in the fourth section, I will argue that, rather than resolving the initial problem, this constitutivist proposal raises the dilemma I have described above. In the fifth section, I will respond to a potential objection to my argument. And in the final section, I will offer a suggestion for how an account of practical reason and rational agency that rejects the guidance condition might nonetheless remain sensitive to the concerns about answerability that motivate the adoption of the guidance condition in the first place.

## 2 The Initial Regress Problem

What is the motivation for imposing the guidance condition on acting for reason? To answer this question, I suggest that we focus on Korsgaard's Kantian-inspired account of acting for reason, where she explicitly defends the adoption of the guidance condition. Korsgaard articulates her account of acting for a reason by distinguishing it from the belief-desire model of action that she associates with David Hume and his adherents (Korsgaard 1997, 220).<sup>2</sup> According to the belief-desire model, a person's behavior is performed for a reason just in case it is caused in the right way by a belief-desire pair of hers. I desire a glass of water; I believe that putting my glass under a running faucet will give me a glass of water; so I put my glass under a running faucet. On this view, the belief and desire, functioning in their characteristic roles, both *cause* and *jointly serve as my reason for my action*.

Korsgaard contends that a belief-desire pair fails to implicate the person's own rationality unless the person herself *recognizes* the belief-desire pair (or its content) to be a reason and unless this recognition motivates her to act.<sup>3</sup> To defend this claim, she has us consider a case in which a person is conditioned to respond to the co-presence of a belief and desire in some incoherent way—say, he

<sup>1</sup> The main aim of Korsgaard's constitutivism is to ground the normativity of practical reason in the constitutive principles of willing or action. But, for Korsgaard, this aim can be satisfied only by showing how agents actively constitute themselves by acting for reasons. Thus, establishing how agents can act for reasons without generating a regress plays an important role in Korsgaard's constitutivist account.

<sup>2</sup> See Davidson (1963) for the canonical contemporary development of this model.

<sup>3</sup> J. David Velleman makes a similar point and proposes a requirement similar to the guidance condition (2000, 9, 194). For reasons on which I cannot elaborate here, however, I do not think his view is vulnerable to the regress problem I describe later.

puts a coin in a pencil sharpener whenever he wants a drink and believes that the object before him is a pencil sharpener (1997, 221).<sup>4</sup> What is the difference, she asks, between this “mad” case and the normal case, in which the belief and desire, and their relation to the action, *are* coherent? A person, she writes, “may be conditioned to do the correct thing ... but the correctness of what she is conditioned to do does not make *her* any more rational” (1997, 221). Again, Korsgaard:

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. A person acts rationally, then, only when her action is the expression of her own mental activity, and not merely the result of the operation of beliefs and desires *in* her. ... Such an agent is *guided* by reason... (1997, 221).

If a person were not motivated by her own recognition of some consideration as a reason, Korsgaard worries, the person herself would not be appropriately, actively involved in the action, and the rationality of the action could not appropriately be attributable to her. A person is *guided* by a reason, on Korsgaard’s view, only if she recognizes some consideration to be a reason, where this recognition motivates her to act. This is what I am calling the *guidance condition* on acting for a reason.<sup>5</sup>

One worry with this picture is that the guidance condition on acting for a reason appears to generate a vicious regress, a worry that Peter Railton raises (2004). Let me give an example to illustrate the alleged problem. Suppose that (A) *I desire greater professional satisfaction* and (B) *I believe that seeking a new job will bring me greater professional satisfaction*. Korsgaard claims that if (A) and (B) simply cause me to seek a new job—which is how the belief-desire model explains the production of an action—then I would be merely blindly conditioned to behave as I do. Furthermore, Korsgaard worries that, if a person were merely blindly conditioned to behave in a certain way, then the rationality of her action could not be attributable to the person herself—for example, we could not appropriately praise, criticize, or hold her answerable or responsible for her behavior. In addition to (A) and (B), she claims, we need something like (C) *I recognize or judge that (A) and (B) constitute a reason*. My seeking a new job is guided by a reason, Korsgaard claims, only if my behavior is motivated by (C).

The problem is that I could simply be “blindly conditioned” to recognize (A) and (B) to be a reason as well. In other words, if (A) and (B) can just “happen to” an agent, then, as Railton remarks, (C) can just “happen to” an agent as well (2004, 185). Think of it this way. I reflect on some consideration—say, the content of a belief-desire pair I have—and I recognize it to constitute a reason for acting. My recognizing it as such motivates me to perform the action. But my recognizing this consideration as a reason might itself be the product of certain other attitudes operating within me, attitudes that motivate me without my recognizing them as reasons. On Korsgaard’s view, these attitudes that motivate me to recognize certain considerations to be reasons are attitudes to which I am blindly conditioned, because they motivate me without my recognizing them as reasons to be motivated or disposed as such. Korsgaard could require that the agent recognize her initial recognition to be a reason, but this sequence could also just “happen to” the agent as well, and so on. The problem, then, is that, by Korsgaard’s

<sup>4</sup> The example comes from Nagel (1970), 33–4.

<sup>5</sup> The last sentence of the quoted passage—“Such an agent is *guided* by reason”—suggests that, for Korsgaard, being motivated by one’s recognition of some consideration as a reason is not only necessary but also sufficient for acting for a reason. It is the necessity claim that primarily concerns me, however. If a proposed necessary condition for acting for a reason cannot be satisfied without generating a vicious regress, then we must either conclude that acting for a reason is impossible or conclude that the proposed condition is not necessary after all.

own assumptions, each reflective judgment a person makes that some consideration constitutes a reason requires rational motivation by some further reflective judgment about the initial judgment, and so on, lest the person be blindly conditioned by some motivating attitude. The guidance condition appears to ensure that acting for a reason could never so much as get off the ground.<sup>6</sup>

### 3 The Appeal to the Constitutive Principles of Action as a Response to the Regress Problem

The regress problem to which the guidance condition gives rise is not simply a problem about the structure or content of justification, considered apart from a person engaged in believing or acting. Rather, the regress problem is about how an agent could possibly satisfy the conditions on *rational self-determination* as captured by the guidance condition. Korsgaard, for one, appears to be sensitive to this worry about the regress. For example, she writes,

The principles of practical reason govern action. Yet I am claiming that reasoning *itself* must be seen as a kind of action, in order to capture the element of self-determination that is essential to volition. If reasoning must be seen as a kind of action, what captures the element of self-determination that is essential to reasoning itself? Do we need some deeper sort of rational activity that in turn captures that? A regress obviously threatens (2008, 84).

What is needed, then, is not simply a reason or principle that somehow halts the regress but a reason or principle *by which the agent actively determines her action* that somehow halts the regress. And Korsgaard argues that the constitutive principles of willing or action as such can fulfill this need.<sup>7</sup>

Constitutive principles are a species of constitutive standards; a constitutive standard of some object describes some feature the object must have in order to be what it is, where this feature is generally determined by the function or purpose the object has (Korsgaard 2008, 8; 2009, 28–9). “If the function of a house is to provide shelter from the weather,” Korsgaard explains, “then it is a constitutive standard for houses that they should be waterproof” (2008, 8). But, according to Korsgaard, constitutive standards are not merely descriptive; they are also normative. A house that fails to keep out the rain is “subject to criticism” for failing to meet the standard that essentially applies to it (Korsgaard 2008, 8). Constitutive principles are constitutive standards applied to activities, and, in the case of goal-directed activities, they “arise from the constitutive standards of the goals to which they are directed” (Korsgaard 2008, 8). Putting one foot in front of the other might be seen as a constitutive principle of walking, just as propelling oneself through water might be seen as a constitutive principle of swimming (Korsgaard 2008, 8–9). Korsgaard’s key move is to explicate the constitutive principles of willing or action as such, the principles that, were we not following them, we would not be acting at all.

How then can these constitutive principles function as regress-stoppers? Recall what gives rise to the regress. According to the guidance condition, a person acts for a reason only if she is motivated by her recognition of some consideration as a reason, lest she be blindly conditioned

<sup>6</sup> See, also, related arguments by Arpaly and Schroeder (2012) for why an act of past, present, or possible deliberation cannot, on pain of regress, be required to act for a reason.

<sup>7</sup> On Korsgaard’s view, these principles include the principle of instrumental reason and, ultimately, the categorical imperative (2009).

to respond to this consideration as she does. But, because her recognition of some consideration as a reason could itself be motivated by some other attitude she has, then it looks like the guidance condition must apply to her act of recognition as well—that is, the guidance condition seems to require that she recognize *her initial recognition of some consideration as reason* itself to be a reason to be motivated as such. But then the guidance condition would apply to that act of recognition too, and so on. Suppose, however, that what guides the person in recognizing some consideration to be a reason is a constitutive principle of action. Take the instrumental principle, which tells me to take the means to my end and which Korsgaard regards as a constitutive principle of willing or acting. Suppose that (A) *I desire greater professional satisfaction*, (B) *I believe that seeking a new job will bring me greater professional satisfaction*, and (C) *I recognize or judge that (A) and (B) constitute a reason*. If (C) is itself motivated by my guidance by the instrumental principle—a principle one follows insofar and only insofar as one acts at all—then it looks like there can be, and need be, no further reason for acting I must recognize in order to act for a reason.

It is central to this account that constitutive principles are both descriptive and normative at the same time. Constitutive principles are normative, Korsgaard explains,

because in performing the activities of which they are the principles, we are guided by them, and yet we can fail to conform to them. But they are also descriptive, because they describe the activities we perform when we are guided by them (2008, 9).

As Korsgaard herself acknowledges, this dual character of constitutive principles seems puzzling. One might object to the coherence of this picture of constitutive principles by claiming that one cannot meaningfully fail to conform to such a principle at all (Korsgaard 2008, 9). The worry is that by failing to conform to a constitutive principle of some activity, a person seems to fail to engage in that activity at all; and it would be inappropriate to criticize a person for failing to meet the standards of some activity she is not engaged in performing in the first place. Consider an example: suppose that propelling oneself through water is a constitutive principle of swimming. And suppose that, during a trip to the pool together, you observe me splashing and flailing in the lap lane, failing to make progress towards the edge of the pool. Because my behavior fails to meet the standards that are constitutive of swimming, it would appear that it does not constitute an act of swimming at all. But if that is right, then it would also appear to be inappropriate for you to criticize me for swimming poorly; after all, I was not engaged in the activity of swimming at all!

Korsgaard's response to this worry is that what is necessary, and also sufficient, for me to meaningfully fail to meet some constitutive standard for an activity is that I am *trying* in some sense to meet it. What is necessary, and also sufficient, is that I am *guided* by the principle, even if I am guided unsuccessfully (2008, 84–5). When I splash, flail, and possibly sink in the lap lane, it is appropriate for you to criticize me for swimming poorly because I am clearly trying to swim well—or at least well enough. My behavior might fail to conform to the constitutive standard for swimming, but I am guided by that principle nonetheless. But while this caveat appears to resolve the worry about failures to meet constitutive standards, I will now argue that it raises a dilemma for this kind of appeal to the constitutive principles of action.

#### 4 A Dilemma for the Appeal to the Constitutive Principles of Action

The appeal to the constitutive principles of action can resolve the initial regress problem only if we have a satisfactory characterization of what following (or trying to follow) a constitutive

principle requires. In other words, we need an explanation of what being *guided* by a constitutive principle of action requires. One conception of guidance is already on the table, namely the conception of guidance represented by the guidance condition: a person acts for, or is guided by, a reason only if she recognizes some consideration to be a reason, where this recognition motivates her to act. Can we apply this same characterization of guidance to following the constitutive principles of action as well?

Suppose that the guidance condition does apply to following a constitutive principle of action. Such a characterization would imply the following:

A person follows, or is guided by, a constitutive principle of action only if she is motivated by her recognition that some course of behavior accords with the principle and by her recognition that this accordance constitutes a reason for so acting.

It is clear, however, that if we apply this conception of guidance—the one represented by the guidance condition—to instances of following constitutive principles of action, then the appeal to these principles does not block the initial regress. Rather, it perpetuates it.

Consider an analogy Korsgaard draws with the principles of logic and evidentiary support. “The principles of logic and the canons of evidence,” she writes,

describe what the thinker as such *does* with the incoming evidence: arriving at a belief through reasoning is an active process, a process by which the mind determines itself to a conclusion. Rational principles may be seen as *directions* in the most literal way. [...] What makes your beliefs logical is not that *they* conform to the rules of logic, for you could believe *P*, *Q*, and *If P then Q*, and never notice any connection between them. [...] What makes your belief logical is that you *put* the two premises together in the way required by modus ponens, and so *cause* yourself to believe it (Korsgaard 2008, 83).

What, then, is required for the agent to put the two premises together in such a way that she causes herself to believe the conclusion? Must she be motivated by her recognition that modus ponens requires her to put the premises together in such a way and that this requirement constitutes a reason to do so? If so, then the possibility would remain that her recognition of modus ponens as a reason was itself influenced by some other attitude operating within her, an attitude to which she would be blindly conditioned given her lack of recognition of it as a reason. To preclude such blind conditioning, it looks like she must be motivated by her recognition of some further principle as a reason, such as *If P*, (*If P then Q*), and (*If (P and (If P then Q))*, then *Q*), then *Q*. Again, regress looms, as Lewis Carroll’s parable of the Tortoise and Achilles famously warns (Carroll 1895).<sup>8</sup>

Suppose, then, that the guidance condition does not apply to following a constitutive principle of action. Indeed, there is some evidence that Korsgaard does not mean to apply the guidance condition to following the constitutive principles of action. She claims, for example, that other philosophers go wrong by thinking of principles as “something like rules that function as deliberative premises” (2008, 227). Rather, she suggests, following a principle is something that is “embodied” in recognizing certain considerations to be reasons (2008, 228). To say that a person follows a principle is not to say that he is motivated by an additional recognition that his proposed action accords with the principle, she argues; rather, it is “just to record the fact that he is active and not merely causally receptive with respect to his perception

<sup>8</sup> I do not mean to suggest that the regress problem Carroll illustrates is generated by a requirement to preclude “blind conditioning” but only that such a requirement—as represented by the guidance condition—generates a structurally similar regress when applied to guidance by principles. For some further discussion of the Tortoise and Achilles regress problem in practical reason, see Railton (1997) and Dreier (2001).

of the good-making properties of the action” (2008, 228). On this suggestion, then, following a principle is something that is *expressed in* a person’s recognizing certain considerations to be reasons, something that is possible even if the person is not motivated by her recognition that her proposed action accords with the principle.

If we accept this alternative characterization of following or being guided by a principle, though, then we appear to face a different problem. Although we avoid the regress problem, we also abandon the initial motivation for adopting the guidance condition on acting for a reason. The requirement that, in order to act for a reason, a person must be motivated by her recognition of some consideration as a reason was meant to capture the way in which a person takes a distinctively active role in determining her action. But this requirement generated a regress. The constitutivist move was an attempt to identify a set of principles *by which the agent actively determines her action* in a way that somehow halts the regress. But it turns out that the constitutive principles of action can halt the regress only if an agent can be guided by them without being motivated by the recognition that her proposed action accords with them. And this suggests that, at bottom, agents can play a distinctively active, self-determining role in their rational lives in the absence of such reflective recognition. But if agents can remain active in the absence of such recognition *with respect to the constitutive principles of action*, then why can they not remain active in the absence of such recognition *with respect to the reasons for which they act* as well?

Thus I conclude that the appeal to the constitutive principles of action, as a way to resolve the initial regress problem, itself raises a dilemma. If being guided by a constitutive principle of action requires that an agent be motivated by her recognition that her proposed action accords with the principle in question—and that this accordance constitutes a reason—then the regress remains unresolved. If being guided by a constitutive principle does not require that an agent be motivated by such recognition, then the rationale for applying the guidance condition to acting for a reason vanishes as well. The way to avoid the dilemma, I submit, is to reject the guidance condition on acting for a reason and thus to accept that a person can act for a reason even if she is not motivated by her recognition that some consideration constitutes a reason.

## 5 A Potential Objection to My Argument: Accepting Asymmetry

The dilemma I have posed for the appeal to the constitutive principles of action issues from the assumption that the guidance condition either applies to *both* acting for a reason *and* following a principle or should apply to neither. It cannot apply to both, on threat of regress. So my suggestion is that it should apply to neither. In other words, if an agent can be guided by a constitutive principle of action without being motivated by the recognition that his proposed action accords with the principle, then likewise he can act for a reason without being motivated by his recognition that some consideration constitutes a reason. After all, if an agent can manifest a distinctive form of active self-determination in the absence of reflective recognition with respect to following constitutive principles, then the rationale for requiring such reflective recognition with respect to acting for a reason vanishes. Rather, it would appear that an agent can manifest a distinctive form of active self-determination in acting for a reason without reflectively recognizing some consideration to be a reason.

One objection to this argument is that accepting an asymmetry between acting for a reason and following a principle (at least one constitutive of action) is simply not so bad. The objection might go as follows: *Acting for a reason is one thing; following a constitutive principle of action is another. The guidance condition applies to the former but not to the latter. Indeed—the objection might continue—it is not clear how the guidance condition could*

apply to a constitutive principle of action; after all, if a person were required to recognize that the principle constitutes a reason, and to be motivated by that recognition, then that person would need to rely on some further principle of action. This would undermine the supposed foundational, constitutive nature of the initial principle.<sup>9</sup> But it is a different story with acting for a reason. Again, if an agent were not guided by her own recognition of some consideration as a reason, then she would be blindly conditioned to do what she does. But this worry about blind conditioning stops where guidance by constitutive principles begins.<sup>10</sup>

To grab the second horn of the dilemma, then, is to reject a univocal account of guidance. And the objection I describe above claims that this implication, far from being devastating, is perfectly acceptable. But this objection fails to acknowledge why such an asymmetry between acting for a reason and following a principle should be troubling. To see why, consider a *skeptical* implication one might draw from the dilemma I have posed. I have concluded that, if an agent can be active in following a principle in the absence of reflective recognition, then she can be just as active in acting for a reason in the absence of such reflective recognition. But one could just as well conclude that, if an agent is merely blindly conditioned to behave as she does, in the absence of reflective recognition, with respect to acting for a reason, then she is just as blindly conditioned, in the absence of reflective recognition, with respect to following a constitutive principle of action.

Another way to see the problem I am posing is to regard it as a worry about accidental conformity to a reason or principle. The guidance condition on acting for a reason is meant to rule out cases in which a person's behavior achieves mere accidental conformity to a reason. This is Korsgaard's point in comparing a "mad" case of acting—illustrated by Nagel's example of a person putting a coin in a pencil sharpener whenever he wants a drink and believes that the object before him is a pencil sharpener—with the "normal" case, in which a person's belief and desire, and their relation to her behavior, are coherent. According to Korsgaard, to establish that the latter agent's behavior is not merely accidentally "rational"—and thus not really rational at all—the agent must be motivated by her recognition of the belief-desire pair as a reason.

But if the worry about accidental conformity applies to acting for reason, then it applies to following a principle as well.<sup>11</sup> Again, suppose that (A) *I desire greater professional satisfaction*, and (B) *I believe that seeking a new job will bring me greater professional satisfaction*. According to Korsgaard, if (A) and (B) merely caused me to act, then, even though (A) and (B) may jointly constitute a reason, it is not a reason *for which* I acted. Rather, I just happened to get lucky in that my behavior accidentally conformed to what is in fact a reason. To preclude accidental conformity, Korsgaard claims, I would need to be motivated by something like (C): *I recognize or judge that (A) and (B) constitute a reason*. Suppose, then, that I consistently judge considerations like (A) and (B) to be reasons to act and that I am then motivated by those judgments. For Korsgaard—according to the asymmetry interpretation at issue—this pattern of judging certain considerations to be reasons is sufficient to show that I am guided by the principle of instrumental reason. But this understanding of guidance by principles fails to acknowledge the possibility that my pattern of judging such considerations to be reasons is itself dumb luck. My judging certain considerations to be reasons might consistently conform to the principle of instrumental reason, but, unless this pattern of judgment is guided by the principle, then—on Korsgaard's own view—such conformity might be merely accidental,

<sup>9</sup> See Katsafanas (2013, 49), who makes a similar point.

<sup>10</sup> Thanks to Adam Leite and Aleksy Tarasenko-Struc for sharing a version of this objection with me.

<sup>11</sup> Notice the conditional character of this claim. I am not claiming here that we should be gripped by the worry about accidental conformity in either case.



mere coincidence. And if an agent's being motivated by her reflective recognition that some consideration constitutes a reason is what excludes accidental conformity in the case of acting for a reason—and thus what makes the agent's behavior genuinely *guided* by a reason—then why think that reflective recognition is not required to exclude accidental conformity in the case of following a principle? Indeed, Korsgaard herself writes, "...if people cannot ever be guilty of violating the instrumental principle *by their own lights*, then it is not a normative principle" (2009, 61, my italics). Thus, although accepting an asymmetry between what is required to act for a reason and what is required to follow a constitutive principle of action may seem promising, the concerns about accidental conformity that motivate the adoption of the guidance condition incline strongly toward a univocal account of guidance, one according to which an agent follows a principle—and thus can violate it—only if he is motivated by his recognition of the principle as a reason. But of course it is this understanding of guidance that generates the regress.<sup>12</sup>

## 6 Conclusion: Rejecting the Guidance Condition but Saving Answerability

If we accept a univocal account of guidance, and if we accept that the guidance condition truly describes a necessary condition for being guided by a reason or principle, then we fall into skepticism about practical reason and rational agency. After all, grabbing the first horn of the dilemma implies that acting for a reason could never so much as get off the ground. Instead, I suggest, we should reject the guidance condition and accept that a person can act for, or be guided by, a reason without being motivated by her recognition that some consideration constitutes a reason. Exactly what this picture of acting for a reason would resemble is an open question, but it is fair to assume that it would follow the spirit of the belief-desire model, according to which the appropriate causal dynamic among one's beliefs, desires, and other attitudes suffices to enable acting for reasons.<sup>13</sup> Taking this avenue immediately raises a question, though: what becomes of the worries about blind conditioning and accidental conformity? Recall that Korsgaard's reason for rejecting the belief-desire model of action is that it fails to include the agent's own mental activity in a way that is necessary for the agent's action to be attributable to her. The guidance condition is supposed to bridge that gap; it is meant to be a condition that, were it not satisfied, the agent's action could not appropriately be attributable to her. If we abandon the guidance condition, must we identify some other condition that excludes blind conditioning or accidental conformity?

One response to this question is simply to reject the assumption that there remains a problem to be solved. According to this line of response, Korsgaard's motivating worries about blind conditioning and accidental conformity are misguided from the start. Railton suggests something along these lines. Perhaps a certain kind of "blindness" imbues our reason-responding capacities, he claims. But, he adds, "The paradox is no deeper than this—a certain blindness enables us to see" (Railton 2004, 190). Furthermore, he suggests, the thought that this kind of blindness undermines attributability by demoting our behavior to mere conditioned responses, whose conformity with reasons is merely accidental, rests on a similar confusion. I, the agent, am not something over and above the dynamic workings of my beliefs, desires, and other attitudes. Rather, Railton argues,

<sup>12</sup> For an enlightening discussion about how some of these concerns about guidance figure in the literature about rule-following—specifically in Wittgenstein (1973) and Kripke (1982)—see Bridges (2014).

<sup>13</sup> See Railton (2004) for an outline of such an account and Arpaly (2006) for a more elaborate development of such an account.

I *am* this self-organized complex that exhibits the proper functioning of thoughts, beliefs, emotions, and desires. The crystallizing of intentions from my beliefs and desires, and the subsequent shaping of my conduct—what I expect, what I am motivated to pursue, what I find rewarding or disappointing—that *is* me in action. The ‘real me’ is no subpart of this or internal manager of it, guiding it from within. The real me—the theory-forming, goal-seeking agent we hope manages to be rational—is the very organism itself. Yet that organism is no mere *bundle* of states. It is a structured, functional whole (2004, 200).

This passage suggests that, if the agent herself is the “structured, functional whole” described above, then the agent’s active involvement in her own rationality—what the guidance condition was meant to capture—need be nothing more than the proper functioning of that structured, functional whole.

The picture of acting for reasons described above presents a plausible alternative to the picture of acting for reasons that adheres to the guidance condition. Indeed I think the worries about blind conditioning dissolve once we acknowledge that the agent is not someone who stands apart from and manages the operation of her attitudes but rather is someone who is constituted by the operation of those attitudes. Because the agent is not someone who stands apart from and manages the operation of her attitudes, it is true that she cannot somehow assure their coherence or rationality (hence Railton’s “the agent we hope manages to be rational”). But of course the guidance condition cannot secure such a status either—or so I have argued. Much of the point of the preceding discussion is that such a self-determining, rational power is not something we can reasonably expect of a rational agent.

Nonetheless, I want to suggest that there is something that the Kantian picture is rightly trying to capture, something to which the alternative picture does not adequately speak. As I have suggested throughout this paper, attributability matters, at least in part, because it grounds judgments of answerability.<sup>14</sup> That is, our interest in determining whether an action is attributable to an agent is explained in part by our interest in whether we can hold the agent answerable for her action. So what exactly is it to be answerable for one’s actions? To be answerable *for* one’s actions is to be answerable *to* criticisms of them, to be responsive to demands for one’s reasons. Andrea Westlund explains that being answerable requires that one “be willing to be engaged in a form of potentially open-ended justificatory dialogue about one’s action-guiding commitments” (2003, 495). Furthermore, we might add, being answerable requires that an agent’s behavior actually be sensitive to her endorsement or rejection of the reasons she articulates in such dialogue. That is, the answer an agent gives in response to criticisms of her actions or demands for her reasons must itself be action-influencing for the agent.

The rationale linking the guidance condition to answerability appears to be something like the following: rejecting the guidance condition implies that, with respect to some of our actions, we act for reasons or principles to which we do not have immediate, conscious access (at least not *qua* reason or principle). And if, with respect to some action, I were unaware of the reason for which I acted, then I would not be able to immediately answer to demands for my reasons. To use a metaphor that has run throughout this discussion, I would be blind to my own reasons on such occasions, and, according to the rationale in question, I would not be answerable for the actions or attitudes those reasons motivate.

<sup>14</sup> Whether the conditions for answerability are identical to the conditions for attributability is a deeper question on which I cannot elaborate in this paper. See Shoemaker (2011) for critique of this identity thesis, a thesis he attributes to Scanlon (1998) and Smith (2005).

I want to suggest, however, that, *ceteris paribus*, I can be answerable to rational appraisals and demands for reasons even on such occasions. What follows is schematic, to be sure, but it is meant to give some shape to an alternative account of practical reason and rational agency, an account that rejects the guidance condition but remains sensitive to the concerns about answerability that inspire the present debate. First, suppose that the appropriate causal dynamic among one's beliefs, desires, and other attitudes—whatever dynamic that might be—suffices to enable acting for reasons. On this suggestion, acting for reasons can be captured by the type of picture Railton gestures at in the passage above. But, second, suppose that it is a separate question of what being answerable requires. I propose that what is required for answerability is not that I have immediate access to the reasons for which I act in virtue of being motivated by my own reflective recognition of them as reasons. What is required, rather, is that, *were I to gain knowledge of the reasons for which I acted, whether immediately or mediately, I would be able either to avow them or to disavow them, where my disavowal would effect appropriate modifications in my future actions and attitudes.*<sup>15</sup> Indeed, it is an advantage of this picture of rational agency that it accommodates the following central feature of our practices: we sometimes hold a person answerable for acting on reasons that, on reflection, he does not endorse. Let me conclude by illustrating how my proposal makes sense of such a case as well as how it makes sense of two varying cases.

Take Robert. For weeks, Robert has been persistently avoiding making his annual doctor's appointment. Although his behavior clearly indicates the source of his persistent avoidance—he fears the prospect of bad news—he never consciously entertains this thought. In fact, his fear likely explains his not entertaining this thought. He gets anxious when his spouse, Sam, reminds him to make his appointment, and he quickly changes the subject. His skin gets clammy when he passes the doctor's office on his way to work. And all of this happens without his recognizing the connection between these behaviors and his avoidance of making the appointment. You see, skin cancer runs in Robert's family, and over the past few years, Robert himself has developed blemishes on his skin that Sam has found troubling. Eventually, Sam confronts him: "You are avoiding the doctor because you don't want to hear bad news about your health!" Robert is indignant at first, but, upon reflection, he realizes that Sam is right. His reason for avoiding the doctor is not that he is too busy, as he's been telling Sam and himself; rather, his reason is that he is afraid of hearing that he has cancer. At that point, he acknowledges that the reason for which he has been avoiding the doctor is not a good one. He picks up the phone, calls his doctor's office, and makes an appointment.

Robert's avoidance behavior is not motivated by his recognition that some consideration—say, his fear of hearing bad news—constitutes a reason. Indeed, his case is an especially clear instance of a person's lacking reflective recognition of the reason for which he acts. (Again, we are assuming that the operation of Robert's fear, perhaps together with other attitudes he has, suffices to make it the reason for which he acts, albeit unconsciously.) The access he gains to his reason is not immediate; it is Sam who reveals it to him. Nonetheless, Sam holds him answerable for his actions. Sam addresses this rational criticism of Robert's actions to Robert himself, with the expectation that he either avow his reason for acting or disavow it and, by doing so, modify his attitudes and actions—the latter of which he ultimately does.

Of course, Robert's case does not represent the more common type of case of acting for a reason without satisfying the guidance condition—one in which, were the agent to gain knowledge of the reason for which she acted, she would avow it unhesitatingly. Take Lilit.

<sup>15</sup> I argue for and elaborate on this thesis in McAninch Andrew (forthcoming) "Activity, Passivity, and Normative Avowal." *Pacific Philosophical Quarterly*. See also Moran's (2001) treatment of the connection between avowal and rational agency.

Lilith and Kat are old friends, but in recent years, as professional obligations have grown for both of them, they've grown more distant. Upon receiving news that Kat has been in an accident, however, Lilith abandons the important project she has been working on and immediately rushes to the hospital to sit by Kat's side. When Kat comes to, she is touched by Lilith's presence—but also slightly surprised. Kat asks Lilith why she dropped everything to be there by her side. Lilith responds, "It's only now that you ask me that I realize how important our friendship is, how much I care about you. That's why I'm here." Indeed, she's right. Of course, Kat's question to Lilith is not offered as a prelude to a rational criticism; nonetheless, her dialogue with Lilith represents the normative expectation that the reason that moved Lilith to act is one she can avow, one that she can articulate as justifying her action. Both Robert's case and Lilith's case suggest that our ordinary practices allow for judgments of answerability even with respect to many agents who do not meet the guidance condition on acting for a reason.<sup>16</sup>

One more type of case is worth discussing—namely one in which, were the agent to gain knowledge of the reason for which he acted, he would (like Robert) disavow it but (unlike Robert) be unable to make appropriate modifications in his future actions and attitudes. Take Frank. Like Robert, Frank has been persistently avoiding making his annual doctor's appointment. And although his behavior clearly indicates the source of his persistent avoidance—he fears the prospect of bad news—he never consciously entertains this thought. After his spouse confronts him, Frank acknowledges that the reason for which he has been avoiding the doctor is not a good one. But instead of picking up the phone and calling to make an appointment, Frank continues to be crippled with fear. Even if he manages to make the appointment, he cannot bring himself to keep it. This pattern continues, all in the face of Frank's explicit disavowal of his fear as a bad reason to avoid the doctor. What should we say about Frank? On the view I am proposing, Frank's fear continues to function as his motivating reason for avoiding the doctor; together with his other attitudes, it fulfills the appropriate functional role to do so. But because Frank's disavowal of his motivating reason fails to effect appropriate modifications of his future actions and attitudes, Frank proves himself to be *not answerable* to rational criticisms of his act of avoiding the doctor. The point is not that Frank fails to exhibit irrationality; clearly he does. The point is that he fails to be answerable to normative appraisals of his action, given that he can neither avow nor *effectively* disavow his motivating reason for acting. Perhaps it would be reasonable for us to demand that Frank take indirect measures to counteract his fear and, in doing so, to *take responsibility* for it in some sense. But once it becomes clear that Frank's fear is wholly impervious to his explicit disavowal of it, then it is not reasonable for us to demand that Frank *answer for* his act of avoiding the doctor. In failing to be answerable, Frank does not fail to act for motivating reasons; but he does fail to express a certain kind of self-determination that we take to be a central element of rational agency.

So I agree with the Kantian view that we should regard the possibility of answerability as a constraint on an account of practical reason and rational agency. And I agree that we should look to our actual practices of holding each other answerable as a guide to developing such an account. But what a description of these practices strongly suggests is that an agent can be answerable even if she is not motivated by her recognition of some consideration as a reason, as long as the reason for which she acts is one she could either avow or disavow, where her disavowal would effect appropriate modifications in her future actions and attitudes. Lilith and Robert meet this condition; Frank does not. Indeed, these cases underscore another affinity that my view has with the Kantian view, despite my view's rejection of the guidance condition. Our reason-responding capacities and our reflective, deliberative capacities do bear an important

<sup>16</sup> Of course, the regress problem implies that no agent can meet the guidance condition on acting for a reason *all the way down*.

relationship to one another. But the relationship is not *constitutive*; that is, the exercise of the latter capacities is not required for the exercise of the former capacities. Rather, the relationship between them is *normative*; that is, rational agents are held to a normative expectation that the reasons for which they act are ones they could avow (or effectively disavow) were they to reflectively, deliberately consider them.

**Acknowledgments** For helpful comments and suggestions, I would like to thank audiences at the University of Pennsylvania, the 2013 APA Central Division meeting in San Francisco, and the 2012 Illinois Philosophical Association meeting in Champaign, IL. I would also like to thank the following individuals: Gary Ebbs, Adam Leite, Kate Abramson, Colin Allen, Matt Carlson, Michael Koss, Susan Blake, Blakely Phillips, Marija Jankovic, Aleksy Tarasenko-Struc, Daniel J. Singer, and Justin Bernstein.

## References

- Arpaly N (2006) Merit, meaning, and human bondage. Princeton University Press, Princeton
- Arpaly N, Schroeder T (2012) Deliberation and acting for reasons. *Philos Rev* 121(2):209–239
- Bridges J (2014) Rule-Following Skepticism, Properly So Called. In: James C, Andrea K (eds) *Varieties of Skepticism: Essays after Kant, Wittgenstein, and Cavell*. De Gruyter, Berlin
- Carroll L (1895) What the tortoise said to Achilles. *Mind* 4:278–280
- Davidson D (1963) Actions, reasons, and causes. *J Philos* 60(23):685–700
- Dreier J (2001) Humean doubts about categorical imperatives. In *varieties of practical reasoning*. Edited by Elijah Millgram. The MIT Press, Cambridge
- Katsafanas P (2013) Agency and the foundations of ethics: Nietzschean constitutivism. Oxford University Press, Oxford
- Korsgaard Christine M (1997) The Normativity of Instrumental Reason. In: Garret C, Berys G (eds) In *Ethics and Practical Reason*. Clarendon Press, Oxford
- Korsgaard CM (2008) The constitution of agency: essays on practical reason and moral psychology. Oxford University Press, Oxford
- Korsgaard Christine M (2009) Self-constitution: agency, identity, and integrity. Oxford University Press, Oxford
- Kripke SA (1982) Wittgenstein on rules and private language: an elementary exposition. Harvard University Press, Cambridge
- McAninch Andrew (forthcoming) Activity, Passivity, and Normative Avowal. *Pacific Philosophical Quarterly*. Moran R (2001) Authority and estrangement: an essay on self-knowledge. Princeton University Press, Princeton
- Nagel T (1970) The possibility of altruism. Princeton University Press, Princeton
- Railton P (1997) On the hypothetical and Non-hypothetical in reasoning about belief and action. In: Garret C, Berys G (eds) In *ethics and practical reason*. Clarendon, Oxford
- Railton P (2004) How to engage reason: the problem of regress. In: Jay Wallace R, Philip P, Samuel S, Michael S (eds) In *reason and value: themes from the moral philosophy of Joseph Raz*. Clarendon; Oxford University Press, Oxford
- Scanlon TM (1998) What We Owe to each other. Harvard University Press, Cambridge
- Shoemaker D (2011) Attributability, answerability, and accountability: toward a wider theory of moral responsibility. *Ethics* 121:602–632
- Smith A (2005) Responsibility for attitudes: activity and passivity in mental life. *Ethics* 115:236–271
- Velleman J David (2000) The possibility of practical reason. Oxford University Press, Oxford
- Westlund AC (2003) Selflessness and responsibility for self: is deference compatible with autonomy?. *The Philos Rev* 112(4):483–523
- Wittgenstein L (1973) *Philosophical investigations*, 3rd edn. Prentice Hall, Upper Saddle