

The Guise of the Good and the Agential Perspective on Action¹

Claire Kirwin
Northwestern University

(This is a draft; please do not circulate without permission.)

1. Introduction

What is happening when an agent acts intentionally? Here is an intuitive line of thought: When an agent acts intentionally, she does the thing she does on purpose. Since she does it on purpose, it must make sense to her to do that thing (here, now, under these circumstances); her action must appear to her as an intelligible object of choice. And for this to be so, she must see her action as *choiceworthy*, and thus as good, at least to some extent and in some respect. When an agent acts intentionally, then, it must be because she sees at least some good in what she does.²

This line of thought, or anyway something like it, gives us the ‘guise of the good’ thesis, a picture of human agency that has held sway among philosophers since at least Plato, who in Book 6 of the *Republic* has Socrates assert that “every soul pursues the good and does whatever it does for its sake” (505e).³ In fact, Plato’s Socrates endorses a particularly extreme version of the thesis, holding that no one ever errs willingly (*Protagoras* 345d-e, *Gorgias* 509e); on this picture, genuine *akrasia* is not possible. For the most part, other proponents of the guise of the good have endorsed weaker variants of the thesis, holding only that the agent must see *some* good in what she does. This is compatible with cases in which the good that the agent sees (the momentary pleasure in the otherwise ill-advised action, say) is acknowledged by the agent to be outweighed, perhaps even overwhelmingly so, by the reasons that speak against the action. Even in these sorts of cases, the basic line of thought outlined above seems persuasive: if the agent was really acting intentionally (rather than being moved by an uncontrollable compulsion say, or perhaps not really understanding what she was up to at all) she must have seen *something* that spoke in favor of the action. Otherwise, why did she do it?

The guise of the good thesis has had a good run. Since Plato, variations on the basic theme can be traced through the work of Aristotle, Aquinas, and Kant, among others.⁴ And in more recent times, the guise of the good has been given a new lease of life through work in action theory by Elizabeth Anscombe and Donald Davidson.⁵ Since then, however, the

¹ Thanks to participants at the 2023 St Louis Conference on Reasons and Rationality for a very helpful discussion of an earlier draft of this paper. Special thanks are also due to Pascal Brixel, Brookes Brown, Kyla Ebels-Duggan, Rory O’Connell, and Quinn White.

² In this paper, I have followed the majority of recent authors in casting the guise of the good as a thesis about whether *the agent* sees any good in her action (Raz, for instance, is explicit that this is the point of the claim). I am inclined to think, however, that the *full* story about intentional action would reveal this focus to be in certain respects misleading. To fully understand intentional action, we have to understand it not just first-personally, but also in an interpersonal context. (See footnotes 15 and 16.) For the purposes of this paper, I shall leave this complication aside.

³ Grube/Reeve. (Strangely, I have also found this passage translated differently—“every soul pursues the good and does its utmost for its sake”—in what purported to be the very same edition of this Grube/Reeve translation.) Compare *Meno* 77b-e for the related claim about desire.

⁴ See for instance Aristotle, *De Anima*, 433a26–30; Aquinas, *Summa Theologica* I–II, Q. 1 Art. 6 and Q. 8 Art. 1; Kant, *Critique of Practical Reason*, 5:59–60.

⁵ (Anscombe 2000); (Davidson 1963).

thesis has also met with an interesting movement of backlash, with Rosalind Hursthouse, Kieran Setiya, Michael Stocker, and David Velleman raising objections.⁶ The basic idea moving the naysayers seems to be something like this. The guise of the good thesis certainly offers a nice, straightforward, intuitive picture of human action. But human action isn't always straightforward, and it isn't always intuitive. It often isn't very nice. People can act for all kinds of reasons, including reasons of malice, hatred, despair, and apathy.⁷ Indeed, the more we think about the on-the-ground reality of agency, the more we are forced to acknowledge that people and their actions can be messed up, complicated, confusing—both to others and to themselves. Nonetheless, the actions in question can be, for all that, still *intentional* actions. (When one makes a malicious comment in order to hurt someone, in full knowledge that it won't make anything better, it needn't be because one is so overtaken by emotion that one simply cannot help oneself.)

This second line of thought seems to me importantly correct. Human action can be pretty radically messy and messed up, and if our picture of agency does not make room for this then it will fail to capture the phenomenon we are interested in. I shall be attempting, in this paper, to offer an account that does justice to this messiness. Nonetheless, I think there is something deeply true and insightful in the guise of the good thesis. Indeed, I shall try to show, in the final section of this paper, that there is something right in the thesis even in its radical Socratic form: there is an important sense in which it is true that *akrasia* is impossible, and that no one errs willingly.

2. What does 'good' mean?

What is really at stake in affirming or denying the guise of the good thesis? In particular, when we say that the agent must see some 'good' in what she does, what does 'good' amount to here? In this section I will outline how I will understand this aspect of the thesis.

Intentional action, Anscombe says, is characterized by the application of "a certain sense of the question 'Why?'". Though Anscombe herself allows that one acceptable answer to this question is "no particular reason", for the most part the interlocutors that I will be engaged with here have tended to think of intentional action in terms of action done *for reasons*—that is, action for which the agent can give a positive answer to the 'Why?' question. I will follow their lead, and largely leave aside cases of action done for no reason, or for "no particular reason".⁸

Sometimes, however, it seems like people can choose to do things precisely because—i.e., for the reason that—those things would be *bad*, in some way or other. Someone may make a cutting remark precisely because it will hurt the recipient, and she may acknowledge in doing so that her reason for acting—that the remark will hurt—does not make her action good at all but instead renders it thoroughly bad. Faced with these sorts of cases, we may want to interpret the agent as acknowledging, specifically, the thoroughgoing

⁶ (Hursthouse 1991); (Setiya 2010, 2007); (Stocker 1979); (Velleman 1992).

⁷ Cf Stocker (1979).

⁸ These two formulations seem to me to point towards different phenomena. A person may smash a plate as an expression of anger, where the action is intentional, but also was not done *in order to* express the anger (thus was not done for an 'expressive reason'), and here she may say: "oh, I didn't do it *for a reason*. I did it because I was angry". By contrast, a person might brush her fingers against a wall as she passes it, again intentionally, but also not in order to achieve something (not in order to learn about the texture of the wall, for instance), and here it makes sense to say, if asked, that she did so "for no particular reason". The former is expressive action, the latter something more like (what has been called) whimsical action—and these seem to me different. Nonetheless, I will leave discussion of such actions for another time.

moral badness of her actions. This need not amount to her using the word “bad” in scare-quotes, as it were, to indicate merely that she recognizes that the action would be conventionally thought bad; she may sincerely judge the action morally bad. But such a judgment is compatible with the agent nonetheless seeing a good of some other kind in the action—perhaps, but not necessarily, a prudential good (I take it that there are many different kinds of good). In any case, it seems clear that the concept of ‘good’ in the guise of the good thesis need not be restricted to the narrowly moral sense of good, and the fact that one may act in ways that one acknowledges as morally bad—or even to do so *because* it would be morally bad—should not be thought to undermine the thesis.

Consider Anscombe’s treatment of the case of Milton’s Satan:

‘Evil be thou my good’ is often thought to be senseless in some way. Now all that concerns us here is that ‘What’s the good of it?’ is something that can be asked until a desirability characterization has been reached and made intelligible. If then the answer to this question at some stage is ‘The good of it is that it’s bad’, this need not be unintelligible; one can go on to say ‘And what’s the good of its being bad?’ to which the answer might be condemnation of good as impotent, slavish, and inglorious. Then the good of making evil my good is my intact liberty in the unsubmitiveness of my will. (Anscombe 2000: 75)

One way of reading this imagined exchange is to see the “good” that is here contrasted with “evil” as, specifically, the moral good. And this allows us to coherently see how there might be something good—in a more expansive sense of good—in doing evil. For one may be after not moral goodness narrowly construed but instead liberty, power, nobility, independence. Thus, as Anscombe has it, although Satan acts in *some* sense ‘under the guise of the bad’—indeed, under the guise of the evil—at the end of the day his action still takes place under the guise of the good as Anscombe understands the thesis.

David Velleman finds such a conception of Satan thoroughly inadequate:

What sort of Satan is this? He is trying to get things right, and so he rejects the good only because he has found respects in which it is unworthy of approval. [...] But then he isn’t really shunning the good, after all, since the goods of liberty and glory remain his ultimate goals. Anscombe’s Satan [...] remains, at heart, a lover of the good and the desirable—a rather sappy Satan. (Velleman 1992: 19)

Anyone who has spent significant time in the company of a person who ruthlessly pursues their own liberty and glory at the expense of all others would probably not share Velleman’s sense of such a character as *sappy*, exactly. But leave that aside. Velleman thinks we ought to be able to make space in our account of agency for a genuinely perverse, a genuinely satanic Satan—for a Satan who pursues *evil*, where this is construed as anti-good, as excluding and opposing *all* forms of goodness, rather than as merely another non-moral sort of good. For this reason, Velleman rejects the guise of the good even on this broad reading of ‘good’.

Nonetheless, towards the end of the paper Velleman makes what might be thought a curious concession. Although he insists that agents acting for reasons need not see their actions as *good*, he does allow that they must see their actions as in some sense *justified* by their reasons:

I cannot act for reasons if I don’t care about doing what’s justified or (as I would prefer to put it) what makes sense. But I can still care about doing what makes sense even if I don’t care about the good. This possibility is demonstrated by my capacity to be guided by what

makes sense in light of a [...] mood such as despair, since what makes sense in light of such a mood just is to do what's bad rather than what's good. (Velleman 1992: 21)

On Velleman's account, action does not take place under the guise of the good, but it does take place, it seems, under what we might call the 'guise of the normative'. The agent acting intentionally must see her action as in some sense justified, warranted, as an appropriate or fitting thing to do given the circumstances (and given her despair, malice, etc), as something that it *makes sense* to do, as something supported in some sense by the considerations that move her. Satan may be evil, but he is not chaotic evil.

Much of Joseph Raz's defense of the guise of the good is targeted against just this sort of position.⁹ Raz argues that we cannot hold on to the normativity of reasons for action—as Velleman seems to want to do—without making use of the concept of goodness; these concepts go hand in hand, if you like (Raz 1999: Ch 2). I think this is probably right. But in fact for my purposes it doesn't much matter, because the 'guise of the normative' that Velleman accepts seems to me to retain enough of the spirit of the guise of the good that I am happy to work with this version of the thesis. As I shall be using the concept, then, the 'good' in 'guise of the good' takes a very broad sense—broader than the sense in which Velleman wants to reject the thesis. You can say that the agent must see some good in what she intentionally does, or you can say that she must see her action as in some way justified or supported by her reasons; for my purposes, what is really at stake in the guise of the good thesis can be expressed in either of these ways.

3. Setiya's challenge

Kieran Setiya agrees. For him, as for me, the question at issue in the guise of the good thesis is really whether the agent must see her action as in any sense supported, justified, warranted by her reasons. Because of this, Setiya finds the above concession from Velleman "perplexing" (Setiya 2007: 37 fn 27)—as Setiya sees it, Velleman had presented himself as arguing against the guise of the good, but then at the last minute he backs away from an all-out rejection of the thesis. Velleman does so, it seems, because he accepts a Razian conception of reasons for action, according to which what it is to be a reason is to stand in a relation of normative support for the action; to act for a reason, then, is to act on the basis of considerations that one takes to stand in this kind of normative relation to one's action.

By contrast, Setiya rejects the guise of the good thesis even in this very broad 'guise of the normative' form. He does so because he rejects the idea, taken more or less for granted by both Raz and Velleman, that reasons for action must be understood in this normatively-inflected way.¹⁰ A person may act for reasons, and thus act intentionally, without seeing her reasons as in any way supporting, warranting, justifying, or 'making sense of' her action—without, we might say, taking her reasons to be in any way *good* reasons.

⁹ See, variously, Raz (1999: Ch 2; 2010; 2011: Ch 4; 2016). Some of the moves Raz makes in these pieces resonate in various ways with the account I develop here—though I am not sure he would at the end of the day accept the picture I end up offering.

¹⁰ Here I shall focus on the argument that Setiya develops in his (2010). In the earlier book (Setiya 2007) he offers an argument that I shall treat as fully separate (although I should acknowledge that my doing so may ultimately misrepresent Setiya's position overall). This earlier argument centers on the point that the guise of the good thesis does not offer an adequate explanation of the distinctive kind of non-observational knowledge that an agent has about her own action (in short, because knowing that something would be good to do is not a way of knowing that you are, in fact, doing it). I am inclined to think that it is a mistake to interpret the guise of the good thesis as aiming to answer the epistemic question so conceived, or at least to do so in the manner that Setiya proposes, but I shall leave this discussion for another occasion.

Setiya's key move is to distinguish two senses of 'reason' in the context of reasons for action. The first is the familiar *normative* sense. In this sense, a reason is a consideration that counts in favor of an action: it is something that supports or warrants or justifies or makes sense of, at least in some respect or to some extent, the action. This is the sense of reason at work when we make claims of the form: p is a reason for A to ϕ . The second sense of reason is a purely *explanatory* one. Here what one specifies is not a consideration that counts in favor of an action, but rather only a consideration on which a given agent actually acted. This is the sense of reason at work when we make claims of the form: A 's reason for ϕ ing (the reason for which she in fact ϕ ed) was p .

Notice, though, that an explanatory reason of this kind still makes essential reference to A 's *taking* p as a reason. To explain A 's ϕ ing by citing, say, her troubled childhood, or the fact that she slept badly the night before, would not be to offer an explanatory reason of the relevant kind—except, of course, in the specific circumstances where A takes those very facts as her reasons (to go to therapy, to take a nap, etc). What we are interested in, that is, are A 's reasons, not just any old psycho-causal story about the origins of her action. In this sense, even explanatory reasons are something to which the agent bears a distinctively first-personal relationship.

The guise of the good theorist will typically want to accept at least some version of the distinction that Setiya relies on. As Raz notes, the guise of the good thesis does not state that agents always pursue things that are *actually* good (Raz 2010: 116). For any number of reasons—because she is mistaken about the non-evaluative facts, say, or because she has confused ideas about what counts as good in this particular context, or in general—an agent may surely act on reasons that do not in fact support the action she does. Her reasons, though they *are* her reasons, and they explain why she did what she did, need not be normative reasons in the sense of actually providing justificatory support for the action. In this way, then, we can indeed get a grip on the idea that we can identify the agent's reason for acting—the explanatory reason—where this does not yet amount to our having identified a reason in the normative sense.

With this distinction in hand, Setiya's central claim is that is sufficient to answer Anscombe's "Why?" question—and thus to mark out an action as intentional—that one offer a reason of the explanatory sort. And since the presence of an explanatory reason does not entail the presence of a normative reason, intentional action can be characterized as such without any reference to normative reasons:

It is sufficient to answer [Anscombe's "Why?" question] that one have a belief of the form "I am doing ϕ because p ," [...]. The object of belief here is a proposition about the explanation of action. That it is cast in the first person cannot alter its logical powers. It does not follow from the fact that I am doing ϕ because p that the fact that p is a [normative] reason for me to ϕ , any more than it follows when those propositions are about someone else. The truth of my answer to the question "Why?" is thus consistent with the absence of any justification for what I am doing, and so I can give that answer without believing, or being required in consistency to believe, that I have such justification. (90)

For Setiya, then, the key point is that intentional action is characterized by the presence of explanatory reasons, not normative reasons, and this is so whether we are looking at action from an external perspective or from the perspective of the agent herself. The agent thus need not see the reasons on which she acts as in any way supporting, justifying, or even

making sense of her action: the guise of the good thesis is false, according to Setiya, even in its very broad ‘guise of the normative’ iteration.

4. First- and third-person perspectives on action

Let us look more closely at the above passage from Setiya. To reiterate, he says:

The object of belief here is a proposition about the explanation of action. That it is cast in the first person cannot alter its logical powers. It does not follow from the fact that I am doing ϕ because p that the fact that p is a [normative] reason for me to ϕ , any more than it follows when those propositions are about someone else.

What exactly is the reasoning here? Here is one way we might try to spell it out: everyone accepts that people in general can act intentionally for reasons that are not in fact good reasons. The presence of an explanatory reason for a person’s action thus does not entail that there is any kind of normative support for the action, and we can understand an action to be intentional in the absence of any such normative support. Since this is true of action in general, I must understand it to be true of my own action in particular (there is nothing that marks *me* out as special, after all). It must thus be possible for me to act intentionally while acknowledging that there is no normative support for what I do.

This line of thought is a tempting one, and seen from certain angles it can appear compelling. After all, if things are true of people in general, then I must surely allow that they are true of me too! But arguments of this form face difficulties, for there are distinctive asymmetries between the things that one can coherently think and say about other people, and indeed about ‘people in general’, and the things that one can coherently think and say about oneself. For instance, consider the following argument, which shares its basic structure with the gloss on Setiya’s argument sketched above:¹¹ everyone accepts that people can believe things that are false. A person’s believing something thus does not entail that the thing believed is true, and we can understand their mental state to be a genuine belief even if it is false. Since this is true of belief in general, I must understand it to be true of my own beliefs in particular (there is nothing that marks *me* out as special after all). It must thus be possible for me to believe something while acknowledging that the thing I believe is false.

There are plenty of interesting philosophical problems on the scene here, concerning how exactly we are to understand Moorean paradoxes, and the issues involved in moving between global admissions of fallibility and the relationship one bears to each of one’s individual beliefs considered in isolation (it seems like it ought to be coherent to hold both that some of my beliefs are false *and*, for each individual belief, that it is not false, for instance). But certainly it seems clear that the argument as given is not sufficient to establish the claim that it is straightforwardly possible to believe that p while acknowledging that p is false.

My point here is not that Setiya himself has made a fallacious argument, for I do not think that this is the right interpretation of his argument. But I do think that reflecting on this tempting interpretation of his phrasing helps to bring to light the point that the move from a third-person perspective on action to a first-person perspective is not a

¹¹ Note that the relevant structure is at quite an abstract level: everyone accepts that people in general can do X... and so on. This means that we do not need to set out the parallel argument in terms of reasons or justification for belief; for the purposes of revealing the difficulties that asymmetries between the first- and third-person raise for the argument, the idea that ‘beliefs can fail to be true’ can substitute for ‘actions can fail to be justified by their reasons’.

straightforward one. There are important asymmetries between the ways we may relate to and understand the actions of others and the ways we may relate to and understand our own actions.¹² Our initial grip on the concept of an explanatory reason, and on the distinction between explanatory and normative reasons, came via reflection on the fact that ‘people’—people in general—are capable of acting intentionally for reasons that are not in fact normative. This observation, while correct, shows us action from the outside. And what the preceding reflections reveal is that it does not immediately *follow* from this observation about people in general that *I* must be capable of acting intentionally while understanding my own reasons to be merely explanatory, and not normative.

5. Alienated intentional action

We cannot infer the possibility of a certain form of self-relation from the possibility of the corresponding form of other-relation. Nonetheless, I think that Setiya is right that this form of self-relation—or at least something like it—is possible. It is possible to be acting for a reason, thus acting intentionally, where that reason does not actually provide any normative support for the action in question, while at the very same time understanding that this is so. This is one way that human action can indeed be thoroughly messed up while still counting as intentional. On this point, I think Setiya is importantly correct. Where Setiya goes wrong, I shall argue, is in concluding on the basis of this possibility that we can understand what intentional action is solely in terms of explanatory reasons, and that there is thus no internal connection between intentional action and normative reasons.

If one understands the guise of the good thesis as providing a necessary condition on something’s counting as an intentional action at all, I am thus siding with Setiya against those who endorse the guise of the good. I do not think, however, that this is the only plausible way of understanding the thesis. My claim will be that although the kinds of cases that both Setiya and I want to make room for do count as instances intentional action, they do so only in a qualified sense. They are not, we might put it, instances of intentional agency in the full-blooded sense. These agents can be seen to act intentionally up to a point, but their doings are intentional actions only in a qualified or defective sense—they are defective, that is, precisely *qua* intentional actions. And this is because the agent’s own relation to the action in these kinds of cases is less than fully agential: she cannot see herself as fully and straightforwardly the *author* of such actions. Or so I shall argue.

It is possible, as we have seen, to characterize someone else’s action as intentional by citing a reason that one doesn’t take to be normative—we may say, for instance: “He spent thousands of dollars on a new business card design in order to impress his horrible Wall Street colleagues. How ridiculous.” In the last section, I noted that this represents a characteristically third-personal perspective on action. Sometimes, however, one does relate to one’s own actions in ways that have echoes of just this kind of third-personal relation: “I’m buying this overpriced nonsense in order to prove to myself that I’m an interesting and tasteful person. How ridiculous.” But precisely because such a self-relation echoes the

¹² To be clear, Setiya does not deny that there are such asymmetries. In introducing his concept of an explanatory reason, I noted that the concept is intended to mark a distinctively first-personal relationship between the agent and her reasons. More specifically, the account that Setiya gives of the way that the agent’s intention incorporates reference to her explanatory reason is such that only the agent herself can bear *this* particular relationship to the explanatory reason (Setiya 2007). A third party may know what the agent’s reasons are, but she cannot bear *this* relationship to them. In what follows, I shall in effect be trying to show that despite this point, Setiya’s account doesn’t fully do justice to the complications that these asymmetries introduce.

characteristically third-person perspective, it makes sense to think of such a self-relation as a form of alienation from the action in question.

It needn't be the case, moreover, that such alienated action takes place beyond the control of the agent. She may be fully capable of ceasing the action or of changing course at any point, and she may be consciously aware that these possibilities are open to her. Further, the action may involve all manner of explicit and careful calculation about ends and means, as well as skillful execution of the parts of the action. Such behavior is clearly very far from a 'mere happening', and it thus makes sense to situate it within the realm of intentional action. But it would be a mistake to characterize such alienated action as intentional without qualification, and if we only had actions of this kind in view, we would not yet have a proper grip on what intentional action really is.

Like Setiya, I do not take my argument against the guise of the good in its necessary-condition form to be primarily a matter of raising counterexamples to the thesis. Any particular purported counterexample can of course be redescribed and reinterpreted so that it no longer conflicts with the thesis, and such back-and-forth between defenders and opponents of the thesis has, Setiya suggests, merely led to stalemate. Instead, my argument is based on a much more abstract claim: that a certain form of self-relation *is* possible, but that where it is present the action counts as alienated and thus as an instance of intentional action only in a privative sense. Nonetheless, in trying to make conceptual room for the peculiar sort of self-relation that I have in mind, it can be useful to think through cases, for the purposes of illustrating what the phenomenon may look like. To that end, I want to think through some of the examples, offered by Setiya and others, that purport to give us instances of agents who see no good in what they intentionally do. The point is not for my interpretation of just *these* cases to do the main load-bearing work within the argument, but rather for this discussion to orient the reader towards the broader kind of phenomenon I have in mind, and hopefully to get to her to see it as something recognizable—whether in herself or in others. In discussing these cases, I shall try to show both (against the traditional guise of the good defenders) that we should indeed be open to seeing these as cases in which the agent sees no good in her action, and (against Setiya) that such actions, though within the realm of intentional action, are not full-blooded instances of that phenomenon.

6. Agents who see no good in their actions

The first case, from Gary Watson (CITE, cited in Setiya CITE), is of a man who thinks that sexual desire is the work of the devil, and thus sees it as providing no reason whatsoever in support of his pursuing sexual activity. It seems easy enough to imagine such a person acting intentionally in order to satisfy that desire. And though it may be tempting to invoke, here, the pleasure that such activity may bring as the good that this man seeing in the action, this would, I think, be a mistake. From his point of view, the fact that it would bring pleasure doesn't make things better at all. For that pleasure comes from the devil, and such pleasure is not good but rather terrible. To do what the devil wants you to do is bad enough; to take pleasure in doing so is surely even worse. And although from *our* point of view, we are unlikely to endorse the constellation of ideas to which this man is committed, our interest here is in how things stand from *his* point of view; whether *he* sees any good in the action.

It would also be a mistake, I think, to suppose that we *must* see such an agent as moved by uncontrollable urges or compulsions. He may reflect, as he undertakes his pursuit of sexual activity, that it is within his power to choose otherwise (perhaps his guilt is intensified by this thought), and I do not think we are forced to assume that he must be wrong about this. So he intentionally pursues sexual activity on the basis of his sexual desire

(that is what he takes as his reason, that is), all while categorically denying that this reason makes the action in any way good.

The second case, from Setiya, concerns an attempt to quit smoking: “When I decide to smoke an entire pack of cigarettes tonight because I won’t be able to do so tomorrow, having quit at midnight, I know that this fact does nothing to justify my action. It says nothing in favor of smoking the cigarettes, which I will not even enjoy.” (CITE) Once again, though for a different reason this time, we should not suppose that the good the agent sees in the action is the pleasure it will bring, for Setiya says the action *won’t* be pleasurable. And, to this reader at least, the possibility of this sort of behavior rings true: though one may do stupid things like this because of the pleasure involved, in many other cases the action *isn’t* actually pleasurable. (Have you ever finished the last few mouthfuls of a rich dessert even while acknowledging to yourself that it doesn’t even taste good anymore?) And again, while such behavior is in some cases rightly characterized as compulsive in a manner that undermines its claim to be at all an instance of intentional action, it needn’t be. One can be perfectly capable of stopping, perfectly aware of this fact, and yet continue on.

In this case, unlike the previous case, we are probably inclined to agree with the agent’s own assessment of how things stand, evaluatively speaking. That he won’t be able to smoke tomorrow, having quit at midnight, is no reason at all—no *normative* reason, that is—to smoke a whole pack of cigarettes now. It *might* be a reason to smoke a smaller number of cigarettes, which will presumably be at least somewhat pleasurable—a reason which, of course, may yet be ultimately outweighed by others (or perhaps not!). But this consideration does not in any way support the action that the agent is actually undertaking. Nonetheless, that is the reason for which he is doing the action.

Faced with such an agent, we will likely feel frustration: “you know full well,” we may say, “that you won’t even enjoy this! You could just flush the whole pack right now—nothing is stopping you! What’s the *point* in doing this when you don’t have to?” These are things we may say to another person. But, of course, they are also things that we can say to ourselves, and in this kind of case we can well imagine the agent feeling a similar kind of frustration and expressing similar thoughts towards himself—doing so, that is, while at one and the same time continuing on with the action. (If you have never had any such experience yourself, I applaud you for your exemplary mental health!)

The final case, again from Setiya, is rather a strange one, for it might seem on the face of it to be exactly the sort of case that is most straightforwardly amenable to the proponent of the guise of the good:

“Why are you running outside in your underwear?”

“Because the house is on fire!”

“What about your family? Won’t they be trapped by the flames? In a circumstance like this, the fact that the house is on fire is a reason to rush upstairs and rescue them, not to look after your own safety while they burn!”

“You’re right. I can’t justify my action at all: the danger is not a reason for me to flee; but it is the reason for which I am doing so.” (Setiya 2010: 90)

This might seem like an odd case for Setiya to choose. After all, someone’s house being on fire seems like just the right sort of thing to serve as a (normative) reason for running

outside. And we might reasonably suppose that an agent who runs outside is likely to have this point in view, even if only implicitly. But Setiya asks us to see this case as an instance of ‘silencing’, in something like the McDowellian sense:¹³ something that might typically be a normative reason has in *these* circumstances (specifically, the circumstances in which the agent’s family remains trapped in the house—and, presumably, the fire is not *so* overwhelming that any attempt at rescue would count as rashness rather than courage, and so on...) no normative force whatsoever. We might initially worry that the case thus relies on quite a contentious thesis about how (normative) reasons work—namely, that they can be silenced in this way, rather than merely outweighed—but I don’t think we need to get hung up on this point. As in the case of the man who thinks sexual desire is the work of the devil, all that really matters is how our imagined agent sees things, and if *he* sincerely believes that the relevant reasons are silenced, this should be enough. (If it helps, imagine the man is—not John McDowell himself, who would surely stay and save his family, but, say, a young acolyte of McDowell’s who is fully committed to the concept of silencing so understood but is, unfortunately, not yet fully virtuous.)

Once again, we needn’t understand this agent to be so overcome by fear that he literally cannot help running from the flames. And though, as in the devil case, *we* might be inclined to endorse his reason as supporting the action (or perhaps not, if we’re inclined to the McDowell-esque picture!), the agent does not see things this way. And so, again, we have a case in which an agent acts intentionally on the basis of a reason which, from his own point of view, does not support the action in question in any respect or to any extent.

I have tried to present these cases in such a way as to reveal why I find them compelling as instances of agents who see no good in what they nonetheless intentionally do. But someone who accepts the guise of the good thesis (at least as it is typically understood) is unlikely to accept my characterizations. Here are some ways she may push back.

First, as I have presented things, the agent in each case believes that her reasons do not support her action. But people are capable of having incompatible beliefs, and so the presence of the belief *that her reasons do not support her action* does not immediately rule out the possibility that she also believes, even if only implicitly, that they in fact they *do* offer (at least some) support for the action. Notice, however, that this second belief will not, in these kinds of cases, look quite like the ordinary case of belief—indeed not even like ordinary cases of inconsistent beliefs. It’s straightforward enough, it seems, for me to hold the following beliefs:

- 1) that I am scheduled to meet a student on Tuesday, August 22nd.
- 2) that tomorrow is Tuesday, August 22nd.
- 3) that I have an empty schedule tomorrow.¹⁴

But the inconsistent beliefs we are considering attributing to our agents in the previous cases are not quite like this. In the meeting-the-student example, I will if questioned be happy to affirm any of these propositions in isolation from one another. But in the cases we’ve been considering, the agent will explicitly deny the proposition that her reasons support her actions in any way—and we may assume, I think, that she is sincere in doing so, at least in the sense that she does not intend to deceive her questioner.

¹³ Setiya has in mind McDowell’s account of virtue in his (1979). I say in ‘something like’ this sense because I am not actually convinced that this particular use of the concept matches exactly what McDowell means.

¹⁴ I believe I got this example from Nic Koziolok.

Perhaps, then, we might want to say that though the agent is sincere in holding that she doesn't take her reasons to support her actions, still *some part of her* must see some good in the action. (We might add that she is in some way self-deceived about this.) I think that often something like this can be a good description of what is happening. But note, first, that whatever the concept of 'parts' of an agent amounts to, such parts are probably not well understood as being mini-agents in their own right, and their 'seeing' things a certain way is not likely to be quite the same phenomenon as an agent's 'seeing'. Further, even if it were the same sort of thing, a *part* of an agent seeing some good in the action is not yet for *the agent* to see some good in the action. (Part of a ship's being blue is not yet for the ship to be blue.)

Of course, these reflections force us to face up to one of the very contentious questions in the guise of the good literature: what exactly is required for an agent to count as 'seeing' some good in her action? So far, I have been working with a gloss on the thesis which has it meaning that the agent takes her reasons to support her action, and this suggests a particularly cognitively robust reading, on which we attribute to the agent a *belief* (whether implicit or explicit). But this is not the only way of interpreting the thesis (see **lots of literature**). I do not have space here to consider this debate in any detail. Here are two brief considerations. First, as has been pointed out in the literature (**CITE**), it seems important that we not end up with a version of the guise of the good on which the thesis is trivial, as it would be if what 'seeing as good' amounted to were nothing over and above the idea that the agent is in fact moved towards the relevant action. Second, the further we move from a cognitively-inflected interpretation of 'seeing', the less the guise of the good still seems to capture the original spirit of the thesis. For instance, suppose we say that an agent has an evaluative experience similar to that of an optical illusion, in which something 'appears' good to her in some sense, but she at the very same time is fully aware of it as, precisely, an illusion. If such an agent were to nonetheless act on the basis of this 'seeing', it is not clear to me that such a case would count as amenable to all proponents of the guise of the good as support for the thesis.

At the end of the day, I am inclined to think that, for the kinds of cases we have been considering, there is something to be said for simply being willing to take these agents at their word, rather than tying ourselves in knots trying to make the cases fit the thesis. These agents *say* they see no good in their actions; perhaps they really just don't.

7. These actions as only qualifiedly intentional

I spent the last section on the offense against the guise of the good thesis, at least as it is typically understood. If we think of the guise of the good as providing a necessary condition on something's counting, at all, as an instance of intentional action, then the sort of psychic possibility that I have attempted to illustrate in the previous section suggests that the thesis is false. In this section and the next, I will try to show why I think there is nonetheless something crucially right in the guise of the good thesis. On my account, the guise of the good does not serve as a necessary condition on something's counting as an intentional action. Rather, it characterizes intentional action's exemplary cases, such that the cases that we have been considering must be understood as, precisely, falling away from the exemplar, and thus as instances of intentional action only in a qualified or defective sense.

Consider again the case of the man who flees his burning home, holding that the fact of the fire gives no justification whatsoever for his action. (One could run through a similar line of argument with any of the other cases.) I have argued that it makes sense to situate the man's behavior within the realm of intentional action, and that furthermore we ought to take

the man's own assessment of his situation at face value, granting that he sees no good in what he intentionally does. To this extent, I agree with Setiya's interpretation of his case. But Setiya thinks that what we have here is an instance of *unqualifiedly* intentional action—perhaps an unfortunate instance in certain respects, and definitely an unvirtuous one, but nonetheless a straightforward and unproblematic exemplar of the type 'intentional action'. This seems to me to be wrong.

The man holds that the reason on which he acts offers no support whatsoever for his action. And faced with this, his imagined interlocutor—the person who first asked him why he was running—has cause, it seems to me, to re-press the “Why?” question: “If the fire *really* strikes you as *no reason whatsoever* to flee... well, *why*, then, are you fleeing?” In response to this question, there are certain answers that will not do for Setiya's purposes: “I couldn't help it—the fear overcame me!” This seems to deny that the behavior was voluntary, and thus undermines its claim to be (at all) an instance of intentional action. “I just ran without thinking—it's only now that I can see that I was wrong to do so. I'm going back in!” Of course unreflective action is perfectly possible, and can surely count as intentional. But, again, this is not how Setiya wants us to understand the case, which was instead supposed to give us an agent who acts fully intentionally while in full awareness of the fact that his reasons are not normative.

The one answer that Setiya's agent can give to the re-pressed “Why?” question is, in fact, the same answer he initially gave: he can repeat the reason that (merely) explains his action. But the sense that there *is*, still, a question to be asked here—the felt need to re-press the question—suggests that the initial answer does not yet suffice. In responding by merely repeating the original answer, the agent appears to refuse the force of the re-pressed question—to deny, in some sense, its applicability. But the re-pressed question is not, I think, a new and different question. It is Anscombe's original question, re-asked precisely because the original answer has been revealed as in some sense inadequate, as not quite what the asker was after. This inadequacy was perhaps temporarily masked in this case by the fact that the explanatory reason given—that the house is on fire—seemed on the face of it to be a perfectly acceptable candidate for a normative reason. But now that the inadequacy of the answer has been revealed, the questioner tries again with the same question.

When, faced with the re-pressed question, the agent can give no answer other than the one already seen to be insufficient, he is thereby failing to treat his own action as a fully-fledged instance of intentional action. His denial of the applicability of the re-pressed question does not take the same form as that of the agent that Anscombe imagines saying “I didn't realize I was doing that”—and this reflects the sense in which his action really *is*, and is recognized by the agent as being, in some sense intentional. But his attempt to refuse the force of the re-pressed question indicates the oddness of the relationship he bears to his own action, and the sense in which he himself understands it as intentional in only a qualified way.

In wanting to respond, again, with the explanatory reason, the agent seems almost to be reporting on his behavior, as though from something akin to a third-person perspective. Certainly his relationship to his action is not actually third-personal, and the oddness of the relationship is in fact a distinctive sort of oddness that can only occur as a form of *self*-relation (you cannot be ‘alienated’—at least not in *this* way—from the actions of another person). Nonetheless, its oddness lies in the way that this self-relation appears to echo the archetypally third-person perspective on action. As such, the agent's position seems to be a strangely passive one. It is as though he *finds himself* or *notices himself* moved by the consideration in question. He does not feel himself to be, we might say, fully ‘inside’ the

action. The action is instead taking place in and through him—indeed, in and through not just his body but also through his very rationality, as he finds himself carefully assessing and calculating about how best to achieve the thing he takes to be not at all good. And again, it needn't be as though he lacks any control over the action, as though his body and mind have been possessed and he is powerless to resist. But still, he does not experience *himself* as the driving force behind the action; he does not take himself to be fully and straightforwardly the *author* of his action. Although the action remains within the sphere of the man's control, such that he *could* stop or change course at any point, still it seems that the action is not in the fullest sense voluntary on his part—he does not experience it as without qualification *his* doing.

I have said that the initial questioner has cause to re-press the “Why?” question. In fact, we have already seen this happening, in a different one of our cases. In the cigarette example, a frustrated interlocuter asked the agent: “What’s the *point* in doing this?”—*why*, that is, are you doing it, given that you think there’s no normative reason to do so? This interlocuter, we can suppose, has already been given the explanatory reason: the agent is smoking a whole pack of cigarette *because* he won’t be able to do so tomorrow. And this answer, it seems, was not enough to defuse the need to ask this version of the “Why?” question. When discussing this case, I noted that we could also well imagine the agent saying the same sorts of things to himself. But this very activity of talking to oneself in this way, of saying, in frustration, just the sorts of things we might imagine a third party saying, highlights once again the strange sense of alienation or detachment between agent and action that we have here. One relates to oneself as though to another person, and as though to another person’s action.

There are at least two ways we might think about the re-pressing of the “Why?” question. The first is as the expression of a certain sort of challenge to the agent (or, in a gentler register, an invitation): to take up the reins of agency, to take ownership of the action, or else to stop doing it. It is a call to stop merely allowing the action to happen and instead to behave *as* an agent. And while one can give oneself a talking-to of this kind, there does seem to be something distinctively powerful about the challenge when posed as a genuinely interpersonal address: *someone else* calls on you to take responsibility for your action.¹⁵ This aspect of the re-pressed “Why?” question is very interesting. It bears heavily on the various forms of responsibility entailed by intentional action—the responsibility one has for one’s action, and the responsibility one has towards others that one be answerable for one’s action and able to offer justifications to those others.

Here, however, I shall leave these interesting considerations aside¹⁶ to focus on another aspect of the re-pressed “Why?” question, namely the way that it is an expression of a certain sort of *bafflement*. We re-ask the “Why?” question because there is something that

¹⁵ For this reason, I am inclined to think that Setiya’s dismissal of what he calls Anscombe’s “habitual appeal to public linguistic behavior” (RWR 26; cf 2010 88-89) is a mistake. Though I cannot do justice to the thought here, I suspect that it is not a mere quirk or strange preoccupation of Anscombe’s that she thinks of the “Why?” question in terms of, precisely, a question to be asked by one person and answered—if an answer is forthcoming—by another. Indeed, although my focus in this paper is on the importance of the agential perspective in action, I think that the *full* story about intentional action must situate within an essentially interpersonal context. Indeed, I am inclined to think that there cannot be such a thing as a genuinely private intentional action, in more or less the same way and for more or less the same reasons that there cannot be a genuinely private language. (Cf Korsgaard, publicity of reasons debate?)

¹⁶ I hope to address them in another paper, “Intentional Action as Interpersonally Constituted,” currently in progress.

strikes us—that strikes both the interlocuter and the agent—as somehow not fully intelligible. But though both interlocuter and agent may experience the action as baffling, I want to suggest that the essential differences between the two perspectives mean that they each take a different form and have a different kind of significance.

When the “Why?” question is re-pressed by someone else, what it expresses is a desire for *understanding*. The lack is primarily an epistemic matter: “Why?” means “Why are you doing it? I don’t *understand*, and what you have told me so far doesn’t allow me to understand.” What the questioner lacks is understanding of the true character of the action before her, and lacking such understanding, she may be inclined to withhold an attribution of intentionality to the action—she may wonder, that is, whether it really can be intentional at all. But the problem here is fundamentally epistemic in the sense that the action’s reality is something that precedes her understanding or lack thereof. She may not understand the action she observes, and she may not be sure whether or not it is really an instance of intentional action. But even if *she* doesn’t grasp its nature, the action may be, for all that, fully intentional.

Things are different when we think about the agent finding his own behavior not fully intelligible. For *him* to find his own behavior baffling is not simply a matter of a failure to grasp the true nature of a thing that nonetheless has its own independent reality. For when it comes to intentional action, the agent’s own self-understanding is at least partly constitutive of the action in question, and of its being a genuinely intentional action at all. To the extent that the agent finds his *own* behavior unintelligible—to the extent that it makes sense to him to say, frustrated, to himself: “Why are you doing this? You *know* there’s literally nothing good about it!”—the very nature of the action as straightforwardly intentional is thereby undermined.

Why must we suppose that what is missing here, what would fix things and render that which is currently baffling instead comprehensible, is the idea of the agent’s seeing some good in the action? Well, one reason to think so is precisely the fact that the merely explanatory reason didn’t seem (*contra* Setiya) to be quite sufficient to answer the “Why?” question, and this suggests that what is needed here is instead something normative, something like the guise of the good.¹⁷ But we must be careful here. If an action and its purported reason strikes us as baffling, as not yet making sense *as* an intentional action, such an action will not necessarily be rendered less baffling if it is pointed out merely that the agent herself took her reasons to be good ones. As Setiya notes,

The following dialogue is futile:

‘She is drinking coffee because she loves Sophocles.’

‘What? That makes no sense at all.’

‘Oh yes it does! She thinks it is a [normative] *reason* to drink coffee.’” (Setiya 2010: 97)

In our previous examples, we had agents acting on reasons that *could*, conceivably, have been normative (recall how we were at first tempted to assume, in the devil case and the cigarette case, that the agent was acting in pursuit of pleasure, where this was plausibly seen as a good), and this gave the actions a certain veneer of intelligibility. But in fact for Setiya, there

¹⁷ Indeed, in Anscombe’s discussion of Milton’s Satan (quoted above in Section 2) she herself puts—what seems to be equivalent to—the “Why?” question as: “What’s the *good* of it?”; what we are after in asking the question, she says, is a “*desirability* characterization,” something that reveals the action as *warranting* desire (both emphases mine). And so it should not surprise us that a merely explanatory reason will not do in answer to the “Why?” question so understood.

are no inherent limits on what kind of thing can count as an explanatory reason for a given action. An agent may in principle take as her (explanatory) reason *anything at all*; she may see no connection whatsoever between her action and her reason—no connection, that is, save the fact that she has indeed so taken this as her reason.¹⁸ And so on Setiya’s account, an agent could (in theory at least—empirical realities about human psychology might make it quite unlikely) drink coffee because she loves Sophocles. Such an action does indeed look very baffling. But Setiya’s point here is that we do not render the unintelligible *more* intelligible by insisting that the agent herself took her reason to be a good one. And this, it seems to me, is correct.

That is how things look from the outside, at least. How do they look from the inside? What if *you* thought that “because I love Sophocles” is a good reason to drink coffee? Would this make your drinking coffee because you love Sophocles less baffling from your own point of view? The thing is, it is very difficult to imagine oneself on the inside of such a thought, precisely because “because I love Sophocles” just *isn’t*—special circumstances aside—a good reason to drink coffee. If you try to imagine thinking that it is, it’s hard to abstract away the knowledge you currently have that it isn’t. But if you *really*, sincerely, did think it was a good reason, and you had a story to tell about why, then I think it would make your action comprehensible from your point of view, and thus constitute it *as* an intentional action. (Consider, if you like, what special circumstances *would* make it a good reason: suppose there’s a coffee drinking competition, and whoever drinks the most gets to choose which ancient tragedian will be immortalized in marble outside the new Classics building... Now your action makes perfect sense, right?) And this suggests that what is lacking in the alienated cases is, as I have claimed, precisely the agent’s seeing some good in the action. They are problematically alienated because and insofar as the agent does not see any good in her action.

Suppose my opponent were to grant this. Perhaps she may yet want to resist my account from a different angle. It may be that these cases do give us instances of intentional action, and that they are in some way defective or alienated where that defectiveness does indeed derive from the fact that the agent sees no good in her action. Still, this does not yet mean that we must see the action as defective precisely *qua* intentional action. Perhaps, that is, these are perfectly decent instances of intentional action, and they are ‘defective’ only relative to some *other* standard. Perhaps what I have been seeking to characterize is not intentional action per se, but rather “unalienated intentional action” or “wholehearted intentional action” or even “virtuous intentional action”, where each of these is conceived of as something over and above intentional action as such.¹⁹

I think this would be the wrong takeaway. For these actions appeared defective, from the agent’s point of view, insofar as she does not feel *herself* fully the driving force behind the action. The agent appears as something more like a vessel through which certain sorts of

¹⁸ I think that to the extent that I accept the concept of ‘explanatory reasons’ conceived of as something different from normative reasons, I cannot get on board with this aspect of Setiya’s account. Explanatory reasons, if they are to do any kind of explaining at all, must rest on *some* sort of pre-existing connection between action and reason, and at the end of the day our only way of making sense of this will I think be via normative concepts, such that ‘merely’ explanatory reasons are ultimately parasitic on normative reasons. But since I have been pushing a different line against his account here, I will leave this particular complaint aside.

¹⁹ Setiya can say something like this, for he holds that intentional action is indeed subject to the normative standards given by ethical character—he just holds that these standards are external to intentional agency itself, so that an action can count as a perfectly legitimate instance of intentional agency even as it falls short of those external standards.

complex mental causes operate. What is lacking is precisely the full sense of *agency*, of its really being she herself doing the thing, and doing it on purpose. And these concepts, it seems to me, concern not matters over and above intentional agency, matters about which we may also be interested for different reasons, but rather get right to the heart of what intentional agency really *is*. If we had in view only cases of the alienated kind, we would not yet have grasped the nature of intentional agency as such.

8. Akrasia

At the beginning of this paper, I mentioned that typical iterations of the guise of the good thesis are compatible, as Socrates' position seems not to be, with the possibility of *akrasia*. The picture that Setiya gives us can certainly allow for *akrasia* too: since it is straightforwardly possible to act for a reason that one does not see as in any way good, it would seem no problem at all to act in service of an acknowledged lesser good over a greater. But I am inclined to think that the very straightforwardness of the picture we have here should trouble us.

An akratic action is, for Setiya, just an ordinary instance of intentional action—a lamentable one, perhaps, but not any more puzzling or less agential for that reason. But akratic action, it seems to me, *is* inherently puzzling. A key part of the experience, which any adequate account must capture, is the sense of agency somehow thwarted—a sense that one experiences while somehow at the same time understanding oneself to be, still, an agent, to be perfectly *capable* of doing what is called for. These two aspects stand in genuine tension with one another, a tension that is inherent in the experience itself.

Notice that in akratic action, too, we have cause to re-press the “Why?” question:²⁰ “granted,” we might say, “you’re doing this for the pleasure it will bring. And you’re right that this pleasure will be a good thing. But you also know that, all things considered, the goodness of the pleasure is seriously outweighed by the other negative consequences; the pleasure just isn’t worth it. And it’s not like you can’t help yourself. So *why on earth* are you still doing it?”. As before, we can imagine this either as an interpersonal address, or as something the agent may say to herself. And again, as before, both the relevance of the repressed question, and the sense of distance that allows the agent to try to reason with herself in this way, both indicate that something is amiss here. Akratic action, like action done for reasons that aren’t conceived by the agent as at all good, does not count as a straightforward instance of intentional action without qualification. Akratic actions count as intentional up to a point, but there is also a sense in which they are defective precisely *qua* intentional action, and if we only had these sorts of actions in view, we would again not yet have intentional action as such fully in view.

This is one way in which I think the guise of the good thesis in its more typical form does not go far enough. In requiring only that the agent see *some* good in her action, it casts akratic action as a totally unproblematic instance of intentional action. It’s not hard to see why guise of the good theorists have wanted to allow this, for ordinary experience appears to show us that *akrasia* is perfectly possible, and furthermore that such action bears a great many of the markers of intentional agency. But when Socrates says that *akrasia* is not possible, he is not saying that this strange but all-too-familiar experience does not happen. He means that a certain conception of *what* is happening—one that attributes *both* full-

²⁰ Thanks to John Brunero for pushing me on this point.

blooded knowledge²¹ of the good *and* full-blooded agency to the agent—will not do. And in this, I think Socrates is probably right. The way to make sense of *akrasia*, as with the kinds of cases I have been considering in this paper, is to see it as a kind of falling-away from agency in its fullest sense. We err all the time, and we often do so while realizing that we are doing so and while realizing that we are capable of preventing ourselves from doing so. But where one knows oneself to be going wrong, there remains an important sense in which it is not one's own full-blooded agency that is steering the ship astray. One's erring is at least not wholeheartedly willed.

Works cited

- Anscombe, G. E. M. 2000. *Intention* (Harvard University Press: Cambridge, MA).
- Aquinas, Thomas. 2006. 'Summa Theologica, Part I–II', Project Gutenberg.
<https://www.gutenberg.org/cache/epub/17897/pg17897.html>.
- Aristotle. 1993. *De Anima, Books II and III* (Oxford University Press: Oxford).
- Davidson, Donald. 1963. 'Action, Reasons, and Causes', *The Journal of Philosophy*, 60: 685–700.
- Hursthouse, Rosalind. 1991. 'Arational Actions', *The Journal of Philosophy*, 88: 57–68.
- Kant, Immanuel. 1996. 'Critique of Practical Reason.' in, translated and edited by Mary J. Gregor, *Practical Philosophy* (Cambridge University Press: Cambridge).
- McDowell, John. 1979. 'Virtue and Reason', *The Monist*, 62: 331–50.
- Plato. 1997a. 'Gorgias.' in John M. Cooper (ed.), *Plato: Complete Works* (Hackett Publishing Company: Indianapolis/Cambridge).
- . 1997b. 'Meno.' in John M. Cooper (ed.), *Plato: Complete Works* (Hackett Publishing Company: Indianapolis/Cambridge).
- . 1997c. 'Protagoras.' in John M. Cooper (ed.), *Plato: Complete Works* (Hackett Publishing Company: Indianapolis/Cambridge).
- . 1997d. 'Republic.' in John M. Cooper (ed.), *Plato: Complete Works* (Hackett Publishing Company: Indianapolis/Cambridge).
- Raz, Joseph. 1999. *Engaging Reason: On the Theory of Value and Action* (Oxford University Press: Oxford).
- . 2010. 'On the Guise of the Good.' in Sergio Tenenbaum (ed.), *Desire, Practical Reason, and the Good* (Oxford University Press: Oxford).
- . 2011. *From Normativity to Responsibility* (Oxford University Press: Oxford, UK).
- . 2016. 'The Guise of the Bad', *Journal of Ethics and Social Philosophy*, 10.
- Setiya, Kieran. 2007. *Reasons Without Rationalism* (Princeton University Press: Princeton).
- . 2010. 'Sympathy for the Devil.' in Sergio Tenenbaum (ed.), *Desire, Practical Reason, and the Good* (Oxford University Press: Oxford).
- Stocker, Michael. 1979. 'Desiring the Bad', *The Journal of Philosophy*, 76: 738–53.
- Velleman, J. David. 1992. 'The Guise of the Good', *Noûs*, 26: 3–26.

²¹ Incidentally, I think that from Socrates' point of view, it matters that we specify 'knowledge' here. Socrates' understanding of the problem of *akrasia* is, I think, ultimately somewhat different from the contemporary casting of the problem in terms only of what the agent *believes* about what she ought to do. But this is a topic for another time.