The New Problem of Akratic Action

Thought & Action in Conflict

I think we should all agree that losing self-control or self-governance is something that mature, sensible, adult human beings do. Or to put this commonplace another way, the loss of self-control—what the Greeks called akrasia and what philosophers translate today as ‘weakness of will’ or ‘incontinence’—characterizes human action. So for example, it belongs to the grammar of akrasia that we can say a person acted akratically. And when we charge an agent with such failure we say that she acted against her better judgment, or that she performed an irrational action.

We can easily see that the failure of akasira characterizes what the agent does when we note that the akratic failure typically persists just so long as the action persists, and that changes in the action incline us to change the sorts of things we say about the failure. So for example, consider the failure of a diabetic on some occasion who akratically eats a piece of chocolate cake. The loss of self-control will persist just as long as her action of eating the chocolate cake persists. She is akratic, in other words, not just at some particular moment, but throughout the eating of the cake. Moreover, the loss of self-control can get progressively worse. She might begin with a nibble, walk to the other side of the room, and then come back and take a very thin slice, only to then return a few moments later, now in almost complete lack of restraint. The increasing lack of self-control in this situation seems to coincide with the unfolding of her action. And finally, that the failure persists just so long as the action persists is clear when we see the agent regain control, in media res, so to speak, after she has already been eating the cake. She
might, after some period of time, see how foolish she is and put a stop to it, throwing the remaining cake into the garbage bin. Here we might say she lost control but came to her senses. When the action comes to a halt, so, too, does her akratic failure.

And yet, in spite of this commonplace, I want to suggest the following, perhaps shocking claim: contemporary philosophical accounts of akratic action have no interest in akratic action. I do not mean to say that philosophers have no interest in akrasia. Indeed, akrasia, the akratic agent, and akratic action have been popular topics among ethicists and philosophers of action during the past five decades. Nonetheless, I maintain that the contemporary philosopher has no interest in what a person does when she loses control. I would wager that to most this claim borders precariously close to the absurd, so my primary aim is to convince you otherwise. Or better yet, to replace the absurd with the outrageous, I argue not only that philosophers have no interest in akratic action, but in fact that they cannot.

A first glance at the literature would immediately suggest that I must be mistaken. Philosophers who work on the various puzzles surrounding akrasia usually begin by specifying that the problem of akrasia is a problem of action. For example, Alfred Mele distinguishes akrasia as a trait of character from akrasia as a sort of action and then rejects the idea that akrasia could simply be a problem of character. He writes that a person might be, in the main, very self-controlled, and yet suffer from a one-off akratic failure. Moreover, akrasia might not infect a person’s character “globally” but rather just with respect to certain areas of her life, e.g. food or sex. Mele then concludes by saying that the topic of his inquiry “is less with the trait of akrasia than with akratic action.” The problem of akratic action, as philosophers like Mele ordinarily conceive it, is thus the problem of figuring out how such actions are possible.
Besides this specification that akratic action is the topic of investigation, philosophers typically offer their explanation of such failure by reference to the action. So according to the standard account—the view held by nearly all contemporary Anglo-North American philosophers—we can characterize the loss of self-governance or control as a break or a gap between what I think and what I do. It is in virtue of this break between my thought and my action that you can charge me with irrationality. My thought about what to do—i.e., my “better judgment”—stands in conflict with my uncontrolled action. So it would seem that action figures prominently into the contemporary picture given that the very characterization of the break makes an essential reference to what I do.

Of course, more details need to be filled in to offer a complete account. The model of the break, as I call it, needs some further specification if we are to move from generalized picture to substantive position. Lots of goings on in our practical lives might be characterized as a break. A break can occur once I form my better judgment but before I start to act. A break can also occur when I suffer a breakdown in the midst of my act. Both cases seem to be what philosophers have in mind when characterizing akrasia as a break. But a break between what I judge I ought to do and what I actually do can also occur through no fault of my own, perhaps because the world will not cooperate. For example, I judge I ought to turn right but have to turn left because the road is blocked. And we might identity a break in the behavior of the addict or compulsive agent who wants to stop what she is doing but cannot. So many species of failure exhibit a break between thought and action, but only some rightly count as akratic failure.

How then should we fill this picture in so that we distinguish akratic failure from these other sorts of breaks? It is at this point that the role of action in the standard account starts to seem mysterious. For although the standard account begins by characterizing a break between
thought and action, the account concludes by specifying a *break internal to thought itself* and so characterizes the failure independently of the action. Just to take a prominent example, Amelie Rorty, who first gave expression to the idea of a break in her essay “Where Does the Akratic Break Take Place?”, distinguishes various “stages on thought’s way to action.” These stages, according to her view, provide “junctures” in which a break can interrupt the movement of thought to action, from inside to out, thus giving rise to the conflict between my judgment and what I do. But in order to explain the conflict, and so to explain the “behavioral akrasia”, as she calls it, Rorty appeals to an internal mental break, or what she calls “psychological akrasia.” So although she began the essay characterizing akrasia as a conflict between judgment and action, Rorty goes on to give an account of the phenomenon in which the role of uncontrolled action is almost non-existent.

Examples like Rorty abound. David Pears offers a similar general characterization by describing a break that occurs in the “escalating transition from thought to action”. At a certain point in that transition when the break occurs, Pears says that the agent makes a “jump” over a barrier thus giving rise to the conflict between thought and action. However, we do not explain this failure by the thought-action conflict, but rather, as he says, by “incorrect processing of information.” Incorrect processing is, of course, a distinctively mental failure. “Irrationality,” Pears asserts, “is a failure to make proper use of material already in the mind.” To take another example, Sebastian Gardener describes this mental break as a “propositional fracture.” And Donald Davidson, the earliest and most influential proponent of the break, writes that “Strictly speaking…the irrationality lies in the inconsistency of the intention with other attitudes and principles rather than in the inconsistency of the action of which it is an intention with those attitudes and principles.” In short, although the standard view wants an account of conflict
between thought and action, it does so only by locating the real engine of failure in the agent’s prior mental break.

This shift from a break between thought and action to a break internal to thought should strike us as peculiar, particularly if what philosophers want to explain is the possibility of akratic action. Here is an account of akratic action that, strictly speaking, doesn’t involve the action at all. Rather than explain the possibility of akratic action, the model of the break locates the failure prior to the action in the antecedent mental states of the agent.

Despite the popular misconception that the standard account can explain akratic failure as a break between thought and action, in what follows I demonstrate why any model of the break on the standard account must view the failure as a break internal to thought itself and so why such accounts must ignore the role of action. This demonstration will come in two stages. First, I show why a certain causal picture must locate the break prior to the action. And second, I show why the attempt to rescue the failure from the prior causes of the action must nonetheless locate the failure solely in relations among thought.

In the next section I establish the parameters for my argument by situating the model of the break with respect to the causal view of action, specifically, the view of action that distinguishes action by prior causes. I then argue that given this causal picture, any such account of the break must conform to two formal specifications if we are to distinguish it from other kinds of breaks. As I hope will become clear, these specifications limit what can count as a satisfactory location for a break and so prevent this model from taking the action into account. I then consider a series of attempts to revive this view which culminate in a rejection of the prior causal picture. The alternative causal conception attempts to see the relevant causes, and so the
relevant break, not as temporally prior to the action but rather as persisting with the action. However, I argue that even this alternative conception must specify the failure of the akratic without reference to the action. If I am right, then at best, action on the standard account merely idles; at worst, action just completely drops out. I conclude with a suggestion for a new problem of akratic action—the problem of giving a philosophical account of what a person does in such irrational failure.

*Action, Causes, and Akrasia*

Before turning to my argument, I want to make obvious the kind of break at issue in the standard account, and how this account depends upon a certain picture of the relation between mind and action.

In some sense, the break should seem innocuous when considered without any additional baggage from a theory of mind and body. According to this model, the break provides the explanation of the akratic failure because it reveals a conflict between what a person thinks and what she does. In the example of the akratic diabetic, the agent eats the chocolate cake despite her judgment that she should refrain, and so her judgment conflicts with her action. This conflict between thought and action provides us with a *rational* break between what she judges she ought to do and what she actually does. Although this rational conflict need not suggest explicit contradiction, it does suggest some sort of inconsistency or incoherence on the part of the agent. And in virtue of that rational conflict we can charge the agent with practical irrationality.

But the canonical story of akratic action does not stop with this apparently innocent notion of a rational break because it also imports a view about the relation between mind and
body. In the wake of Davidson’s early work, philosophers typically think of the relation between thought and action as a \textit{causal} relation. Thought stands to action as cause stands to effect. Moreover, on Davidson’s picture, for something to even count as an intentional action it must be caused in the right way by some relevant thought of the agent. So not only does thought stand to action as cause stands to effect, but in order to have any intentional action, some relevant thought \textit{must} cause the action.

This understanding of thought’s relation to action has consequences for how we understand the akratic break. To see why, I want to make the contours of this causal view a bit more precise. Specifically, we need to be clear on what sort of thing an action is and what the relevant sort of thought is.

First, according to Davidson’s view, an action is an event that is both datable and particular. The eating of the chocolate cake by our akratic diabetic is a particular happening that takes place on a given day at a given time. And furthermore, on this story, an action is an event that is identical with the movement of a body. So for example, the event which is her action of eating the chocolate cake will consist in the movements of her body as she eats the cake, i.e., the movement of limbs and mouth as she cuts, lifts, and chews. So actions are physical, particular events that take place at a given time.

Of course, not just any event of an agent moving her body will count as an intentional action. Sometimes our bodies move, but not because we move them, as when my hand twitches or I jerk because of a startle response. Or to take the more common example from the literature, my arm could rise because I raise it, or it might rise for some other reason, as when it rises into the air from a spasm. On the standard approach, since action might appear indistinguishable from
a mere movement of a body, we need to identify some difference between them in virtue of which the one counts as an intentional action and the other not. That difference, argues Davidson, can be found in the distinct causal histories of the two movements.

Second, Davidson identifies the relevant sort of thought which causes the action with some attitudes or mental states of the agent which rationalize the action. These attitudes and mental states, a belief-desire pair on Davidson’s view, provide us with the reason the agent acted, for an action, unlike a mere happening, is subject to explanation by reasons. The reason why the agent acted reveals what the agent saw in the action, i.e. how the action was favorable, desirable, or appropriate by the agent’s lights. Thus, on Davidson’s view, the relevant sort of thought provides us with a logical connection between that thought and the action.

But the thought which provides the reason cannot merely explain the action. After all, the agent could have several distinct thoughts, each one of which rationalizes the action or shows the favorable light in which the agent saw the action. But a genuine reasons explanation does not merely cite one of several possible thoughts. Rather, a reasons explanation cites the reason why the agent acted. That is to say, where some thought explains the intentional movement of the agent’s body, we say that the agent acted because of that reason. This ‘because’, according to Davidson, expresses the causal relation between the thought and what the person does.

The relation, then, between thought and action is both rational and causal. Or to use Davidson’s idiom, “a reason is a rational cause.” An action is a datable, particular event, the movement of the agent’s body. The relevant thought, in virtue of which the movement counts as an intentional action, is some attitude which rationalizes the action and causes it to come about.
Given this picture of mind and action, the standard story of akrasia cannot identify a break that is only rational, for any rational break in the relation between thought and action must also be a causal one. The akratic break, in other words, does not merely pick out a rational conflict between what I think and what I do. Rather, the break will pick out a temporal moment between the physical, datable, particular action and the mental antecedents which cause the action.

**The Possibility of a Break**

I begin my argument now with the model of the break characterized in the most general terms, i.e. as a break between thought and action, without specifying what sorts of thought might be involved. What inferences can we draw from this model in its most general terms?

Suppose that we represent the progress of thought and action in the standard account pictorially along a line. (See Figure 1) Let the direction of the line from left to right indicate the movement of time. Given the direction of progress and the assumptions of the prior causal account, the section of the line on the left will indicate the antecedent mental activity of the agent. According to the standard account, this section would include the agent’s deliberations, her practical reasoning, decision, choice, intention and so on. In short, the left side of the line represents what Rorty calls the ‘stages on thought’s way to action.’ The right side, by contrast, will represent the unfolding progress of the action. For the standard account, this side consists, in the main, of the motor movements of a human body. Together, the complete line represents what Pears calls the ‘escalating transition from thought to action’.
Now, if we begin with just this level of generality, we can easily represent the characterization of akrasia offered by the standard account. To do this, we might simply amend our diagram in order to show the gap that appears between the two sides of the line. However, we could also draw such a diagram in order to represent the breakdown when the world fails to cooperate or when I pursue my addiction against my will. So we need to represent something further on the diagram if we are to adequately distinguish the akratic break. In order to make this distinction and so to represent the standard account, I suggest that any possible account of the break must satisfy two fundamental conditions: 1) first, a specification on what comes prior to the break and 2) second, a specification on what comes after the break. These specifications, I argue, reveal that any guise of the general model must locate the failure independently of the action, and as such, the action will play no role in the account of akratic failure.

1. Prior to the Break. The first specification restricts the sort of judgment that comes prior to the break. In order to generate a rational break between thought and action, the agent cannot act contrary to just any thought which happens to pass her mind, but only one which in some way determines or settles the question of what is to be done. For example, I would not count as akratic were I to idly remember that I need to change the oil in my car before a road trip. Nor is it generally considered a rational failure when someone passes by a homeless man on the street and wishes that she did more for charity. These thoughts would not even generate akrasia if they resulted during some episode of reasoning. For example, I could prioritize all the things I need to do before I leave on my road trip, and as the result of those considerations conclude that I should ideally change the oil before I go. Nonetheless, I could reach this conclusion all the while knowing that I have no time or intention of going to the mechanic until after I get back. I consider this thought for a moment as the result of my deliberation and then blithely continue to
review my pre-travel checklist. So idle thought and wish cannot settle the question on behalf of some course of action, nor will just any thought we happen to infer during deliberation.

It would seem that for any guise of the break we need to specify what sort of thought settles the matter in favor of one particular course of action such that what I now do is rationally constrained. The relevant thought or judgment, in other words, needs to be one whose logical features govern what I do. This idea of a thought which rationally constrains my action suggests a kind of strength in the connection between my judgment and deed, a strength that mere idle thought lacks. According to the standard account, we can identify one’s concluding ‘better judgment’ as the ideal candidate to play this role.19

But even if we suppose this view correct and so attribute a special strength to our better judgment, one must also keep the rational linkage between thought and action loose enough such that akratic action is still possible. Consider, for example, the negative thesis of Socrates and Hare which denies the possibility of genuine akrasia. According to Socrates’ account in the *Protagoras*, practical thought and action cannot come apart because the positive evaluation in one’s better judgment guarantees that the agent will always have the relevant motivation.20 In a similar line of thought, Hare argues that it belongs to the logic of evaluative language such that if I judge that “I ought to do x”, then I also give assent to the command, “Let me do x”. And, according to Hare, it is impossible to assent to a command that one should act thus-and-so and yet act contrary to it.21 In other words, the very judgment that rationally constrains one’s action also makes it impossible for one to act otherwise.

The negative thesis of Socrates and Hare cannot be maintained. As almost any mature, adult human being can attest, akratic failure is an unfortunately common experience in our
practical lives. But if we reject their thesis, then we cannot construe the rational linkage between what I think and what I do to be so tight. Thought does not guarantee action. Rather, we need to allow distance between my better judgment and my action so that akratic action is still possible. Moreover, when I succeed and overcome obstacles in order to do what I judge best, we need that distance in order to make sense of the exercise of self-control.

We have then two limitations on the judgment which comes prior to the break. First, the problem of idle thought demonstrates that we need a judgment whose features or properties rationally constrain what the agent does. Second, by rejecting the negative thesis of Socrates and Hare we see that the connection between judgment and action cannot be so tight as to rule out the possibility of akratic action. In other words, the judgment prior to the break needs to have the power to rationally constrain what I do without thereby guaranteeing that I will act thus-and-so.

This first specification of the break marks a first step towards an adequate representation of the general model. (See Figure 2) According to this line of thought, a concluding or final better judgment has the power to constrain action without thereby guaranteeing action. One’s final judgment during an episode of deliberation should have the strength which mere idle judgment lacks. And the finality of the judgment further suggests that it offers our best, considered opinion on the matter as opposed to just any conclusion we might happen to reach during the process of deliberation. Supposing then that this better judgment concludes the process of thought, we can amend our diagram, locating the break at the immediate end of the agent’s deliberation and just prior to her action. However, this representation remains very much incomplete for we still do not have adequate means to distinguish the akratic from other possible breakdowns.
2. After the Break. The second specification restricts the moment that comes just after the break, i.e. the item which follows the agent’s better judgment. According to the standard characterization of akratic failure, the break comes between my thought and deed, as evinced by our newly altered diagram. But understood as the temporal progression of thought to action, this diagram misleads us. In order to keep intentional action in view, it cannot be the case that an akratic action immediately arises following one’s better judgment. That is to say, once we have specified the thought prior to the break such that the agent’s possibilities for acting well have been constrained, it cannot be the case that the agent just straightaway acts in a contrary fashion. According to the causal conception of action in play, any action whether controlled or uncontrolled must arise from some antecedent mental state in order to even count as action, i.e. in order to recognize what is done as an exercise of agency. But on the model of the break, taken simply as a gap between some prior action-settling thought and the akratic act, it appears as though the action happens in a vacuum, appearing unheralded out of nowhere. Indeed, it might seem as though the uncontrolled act just besets the agent out-of-the-blue following her concluding better judgment about what she ought to do. Such activity, while no doubt frustrating to our practical goings-on, would have little in common with the phenomenon of akratic action. Indeed, given that the behavior would simply confront the agent out of nowhere, it would have a much great similarity to compulsive acts, nervous ticks, or startle responses. Genuine akratic behavior, although it might feel like an assault on our agency and although we might describe it so on some occasions, must nonetheless arise from the exercise of the individual’s own agency and as such cannot simply beset the agent following her concluding better judgment.

Thus, in order for any guise of the break to avoid the appearance of out-of-the-blue akrasia, the item which immediately follows the break cannot be the akratic action itself. Rather,
our second specification suggests that we need some further thought or judgment in virtue of which the akratic action counts as (properly authored) intentional action done for a reason. The Davidsonian causal view holds that some motor movement counts as an intentional action done for a reason in virtue of the antecedent mental states which both rationalize the action and cause it to come about. Intentional action, in other words, whether rational or not, must stand in rational and causal relations to thought. And if akratic action still stands in such relations to thought in order to count as intentional action, then the model of the break must allow some sort of thought or other mental state to appear on the same side of the break together with the akratic action. In other words, given that akratic action is still action and not compulsive movement or some other mere happening, then the moment just after the break will be some further action-guiding thought rather than the action itself.

The second specification thus adds a further limitation on what can count as a legitimate representation of the break. The worry that one’s behavior might become unhinged from our capacity as agents demonstrates that we cannot locate the break immediately following all activity of thought. Rather, we must still allow some additional thought to follow in virtue of which the akratic action will count as action. (See Figure 3) This specification thus allows us to amend our generic diagram, pushing the location of the break temporally and causally prior to the scene of the action. Of course, as with the first specification, not just any thought which follows the break will have this capacity. It could not be a mere idle thought, nor could it be a thought which impugns the rational constraints of the prior judgment. And although, like the first specification, this additional requirement leaves the model of the break in relatively generally terms, we now have a relative location that situates the fault between two specific moments.
The (Reappearing) Mental Break

We have now seen why adherents of the break not only do not, but cannot have any interest in what a person does when she suffers from akrasia. These two specifications on the model arise purely from the general form of the model itself, together with the traditionally accepted grammar of akritic action. From these few details, we can see why Pears considers irrationality as “incorrect processing” in the mind, and why Davidson sees it as mere internal inconsistency. The action cannot figure in the explanation of the failure because the failure will already have taken place by the time the agent begins to act.

How can a proponent of the break respond to these specifications in order to get action back into the picture? And what must such an account look like in order to keep the loss of self-control from reducing to a mental fracture? It seems implausible to reject the first specification for, *ex hypothesi*, the akrates acts against her better judgment. If we abandon the first specification then the standard account would lose its grasp on the idea that the agent’s behavior is rationally constrained by some prior thought or judgment. And to lose that rational constraint is just to give up on the idea that we have akritic actions at all.

A more plausible response would be to reject the second specification as I characterized it. To see how this response might go and to keep things clear, let’s call the action-guiding thought that follows the akritic break the agent’s ‘akritic thought’. As I imagine this response, an adherent of the standard approach might argue that even though an akritic thought must always follow the break, we need not suppose that this akritic thought thereby guarantees any particular action. For example, I might judge that writing my paper is the best thing to do, but then have an akritic thought to go for a swim instead. If my akritic thought *guarantees* that I
will act contrary to my better judgment, then my going for a swim is actually an unnecessary extra component in our account of the failure. The failure, for all practical purposes, has already taken place the moment I have that thought. However, this response asks, why should the akratic thought provide that guarantee? It seems plausible to suppose that after the akratic thought arises that I could still regain control and act in accordance with the original better judgment. In other words, just as my better judgment doesn’t guarantee continent action, my akratic thought doesn’t guarantee akratic action. And if akratic thought doesn’t guarantee akratic action because the agent could exercise self-control, then we can identify a second possible break between the akratic thought and the akratic act. This further possibility would then allow the model of the break to bring the akratic act back into the picture, because we would not know until the agent begins to act whether she will act akratically or whether she will exercise self-control. So as long as we can make sense of that second possible break, then what a person does could not be a mere additional byproduct or idle wheel. The possibility of exercising self-control means that we need to keep the action in view in order to tell the complete story.

Unfortunately, this response will not do, for the possibility of a second break will only repeat the conditions that led us to the problem in the first place. Consider the classic formulation of the mental break by Davidson.24 (See Figure 4) According to his account, the akratic break arises between two different evaluative judgments. Prior to the break we have the agent’s better judgment, what he calls an all-things-considered evaluative judgment (ATC), which settles the question of what is to be done. After the break we have the agent’s concluding judgment, what he calls an all-out evaluative judgment. According to Davidson, the ATC judgment allows for the possibility of a break because it is only conditional or provisional in form. By contrast, we cannot fail to act on an all-out judgment because it is unconditional and final and so issues
directly in action (so long as nothing prevents us). When all goes well, the continent agent forms her all-out unconditional judgments in accordance with her ATC judgment. But in the akratic case, the agent makes an irrational transition from her ATC judgment to an akratic unconditional judgment.

Now suppose we take seriously the response to my claim and so modify the Davidsonian framework. The modified picture would allow for a second break following the agent’s unconditional evaluative judgment. So contrary to Davidson’s original idea, this means that an agent who suffers weakness of will following her ATC judgment—i.e., an agent who has an akratic unconditional judgment—could still exercise self-control. The problem, however, is that if our agent need not act on her unconditional judgment in the akratic case, she need not act on it in the continent case either. The move which creates the possibility of a second break simply removes the action guaranteeing force of the agent’s unconditional judgment altogether, for it would be arbitrary to assert that the unconditional judgment guarantees action in one case but not the other. But if an akratic break can arise following our unconditional judgment, then in keeping with the second specification, we need to introduce some further thought following that judgment in order to keep intentional action in view. And of course, by introducing some new thought, we shift the akratic fault back between two mental states and thus once again render the action an idle wheel.

To see how the mental break might specifically reappear, let’s look at Michael Bratman. In his early criticism of Davidson, Bratman recognizes the possibility of this break following an agent’s unconditional evaluative judgment. He argues that Davidson was mistaken to assume that the break had to arise in the midst of our evaluative reasoning, since both the conditional and the unconditional evaluative judgments seem apt to the play the role of our better judgment.25
(See Figure 5) According to his alternative account, we should rather locate the break between the agent’s evaluative and non-evaluative judgments. All evaluative judgments take the moment prior to the break and non-evaluative judgment of the form “I shall φ” takes the moment after. In short, the characterization of akrasia as a mental break appears once again.

In order to bring the action back into Bratman’s picture we could reintroduce the possibility of self-control, this time following the akratic agent’s non-evaluative judgment. But once again, this new break following the akratic non-evaluative judgment produces the same possibility following the continent non-evaluative judgment. And indeed, Pears argues contra Bratman that we must allow a break after our non-evaluative judgment since Bratman identifies the concluding non-evaluative judgment with the agent’s intention.26 (See Figure 6) Pears suggests that on a natural understanding of the phenomenon, mature adult human beings act contrary to what they intend all the time—i.e., not merely against what they judge best to do, but also against what they intend to do. So even if Bratman is right to identify a possible break between evaluative and non-evaluative judgment, we must still allow for an akratic break following intention if Pears is correct. But as you might assume, this new possible location for the akratic break will not suffice to get action back into view. On Pears’ account we posit yet another thought following the break, this time a further intention to act thus-and-so. Thus, according to Pears, akrasia arises between our inner and outer intentions and so it still arises between two thoughts.27

So now we try our rescue move again, this time suggesting that the akratic need not act on her akratic outer intention. Were she to exercise self-control, we might suppose, then we could keep the action as part of the explanation. And yet, just as our previous two attempts failed, here again we confront the same problem. If a break can arise following an akratic outer
intention, then surely it could arise following a continent outer intention. And in acknowledging a break between intention and action, then the second specification will push us to once again posit some further mental item that comes between our intention and what we do. But what might come after intention? And what further mental state could we add that would possibly stop this cycle? Surely any mental state we add will create yet another possible location for a break. If that is the case, then we must either give up the second specification and so allow what I call out-of-the-blue akrasia, or else we must accept the unfortunate conclusion that the model of the break will never be able to account for what the akrates does.

One possible alternative remains if we want to hold on to the break. Alfred Mele agrees that we will not find some further mental state to explain the failure and to account for the exercise of our agency in akratic action. So instead of positing yet another mental state, Mele suggests that the real way out comes through the exercise of self-control.\(^{28}\) (See Figure 7) In each of the previous rescue attempts, we suggested that a break might arise if the agent could exercise self-control. But rather than use that exercise in our explanation, we appealed to yet another mental state. Mele does just the opposite. He suggests that the break is not between our intention and something further thought, but rather between our intention and an inner act of self-control. Rather than posit a mental state, Mele suggests that the agent herself can intervene through this inner act. If the agent is successful, she will act continently by exercising self-control and so shifting the balance of motivation back to her better judgment. If she fails, either because she doesn’t exercise self-control or because her attempt at control was inefficient, then the akratic thought prevails and so she acts akratically. Mele concludes that we should think of akrasia on this picture as a failure to exercise self-control rather than acting against one’s better judgment.
With Mele our rescue attempt comes to a disastrous conclusion. Although his approach has the virtue of apparently halting our ever expanding botanic garden of mental states, by introducing inner acts Mele in fact introduces a second act and so introduces an entire new range of potential breaks to contend with. Now we have an inner act as well as an outer act. And just like our outer act, this inner act must presumably stand in rational relations to thought. Of course, if this inner act of the agent does stand in rational relation to our prior thoughts, then any possible break that arises between thought and the outer act could also arise between thought and our inner act. I might, for example, intend to exercise self control but fail to do so when the time for acting arrives. Adding further thoughts won’t help us any more than they did with the outer act. Since that approach seems doomed to failure, we might instead try to repeat Mele’s move and so avoid positing such further thoughts altogether. But if we take this approach we will need a second inner act—an inner-inner act of the agent which brings about the inner-act, and thereby brings about the outer act. It would seem that in following Mele either way, a multiplicity of breaks looms on the horizon and thereby an unacceptable regress.

*Deviant Akrasia*

As I see it, adherents of the standard account face a decision. Either they need to hang on to both specifications and so just accept that akrasia is a mental failure rather than a failure of action, or else they can abandon the second specification, and thereby lose their grasp on the idea that akrasia counts as intentional action. In either case, it is not really clear why we are still talking about action. With the former, action appears as a mere byproduct of my failure and so simply idles. With the latter, action seems to completely drop out.
At this point a defender of the standard account would insist that I have simply misconstrued the causal theory of action and thus presented a false dilemma. Someone following the Davidsonian causal view might argue that action, as I put it on the first horn of the dilemma, is not a mere byproduct of the antecedent psychic states of the agent. Rather, action just is defined as an event that has a particular kind of mental cause. According to this view, what makes the rising of my arm an action of raising my arm cannot be explained except by appeal to such states. So if we want to keep action in view—whether akratic or not—then we need to understand the causal connection between the relevant mental states and the movement of the agent’s body. Or to put it another way, the action does not idle on the model of the break precisely because the causal relation between those antecedent mental states and the agent’s behavior is constitutive of what it is to be an action.

However, insisting on the constitutive character of this causal relation offers a hopeless strategy to rescue the standard approach since the causal theory of action runs into notorious difficulties precisely on the issue of whether or not it has intentional actions in view. The most problematic counterexamples start from instances where the causal relation obtains between the relevant mental states and some apparent action rationalized by those mental states and then points out that the causal relation is in some way deviant or wayward. For example, Davidson himself tells the story of a pair of climbers on the side of a mountain. Half-way up the mountain one climber has the desire to be rid of the extra weight on the rope and he has the belief that were he to let go of the rope, thereby dropping the other climber, that he would be free of that extra weight. These thoughts so unnerve our climber that he subsequently let’s go of the rope and so drops his companion. In this case, it would seem that the causal relation holds between the agent’s thoughts and the movement of his body in precisely the way required to produce
intentional action. The climber’s beliefs and desires rationalize the climber’s letting go of the rope. And furthermore, it is just these thoughts which cause the movement of the climber’s body, i.e. the movement of his letting go of the rope. And yet, in spite of this causal connection between the antecedent mental state and the movement of the body, we do not have intentional action in view because the climber’s movement does not count as an action.

Insisting on the causal relation between thought and bodily movement as an account of action simply cannot rescue the model of the break. The problem of deviant causal chains suggests that the presence of such causal connections will not suffice to distinguish intentional actions from mere body movement. And in fact, given the possibility of deviant causation, it will be no surprise that the model of the break can generate climber-style counterexamples of its own, i.e. situations in which the akratic thought causes an apparent uncontrolled act and situations in which an attempt at self-control causes an apparent self-controlled act.

First, let’s consider causal deviance when an agent appears to lose self-control. Imagine an agent considering whether or not to go to the pub for a drink as he drives down the road. Suppose that he needs to finish his term paper and he knows that if he goes to the pub he will not finish in time. Further, let’s suppose that he also knows that he is constitutionally weak in certain situations. In this case, he knows that if drives past the pub that he cannot resist going in because the lure of drink will be too strong. Unfortunately, his normal route home takes him past the pub which will soon be in view only a few blocks ahead. If he turns his car at the approaching intersection, he could take an alternative route home and so avoid driving past the pub. If he keeps driving straight on, he will end up in the very position he wants to avoid. He judges then, all things considered, it would be better for him to turn his car at the approaching intersection. Yet in spite of his better judgment, he then has the akratic thought to give into his desire for a
drink so to keep driving. If he gives into this thought and drives straight ahead, he will knowingly act against his better judgment. If he resists and turns at the intersection, he will exercise self-control. Now imagine that this driver, so focused as he is on his desire to keep driving, misses the turn at the intersection and so finds himself in front of the pub after all. The driver seems to do the very thing he judged he shouldn’t do, namely, keep driving towards the pub. Moreover, there seems to be a causal connection between his akratic thought and what he does. And yet, I think we should say that he does not act akratically. Although it is true that his thought causes what he does, we do not have intentional action.

Second, let’s consider a climber-style case of self-control. Suppose our diabetic stands on a balcony, holding a plate with a piece of chocolate cake. Suppose that she really wants to eat it, but knows that she should not. After an episode of deliberation, she decides that she should simply toss the cake over the edge of the balcony. She could set the cake down on the table, but she is afraid that she might give in as long as it is available. The only way to exercise self-control, she concludes, is simply to be rid of the cake once and for all. However, in considering the cake and her desire for self-control, suppose that she startles herself by the strength of her resolve since in the past she has often given into chocolate temptations. As a result, her hand trembles and the cake slides off the plate and over the balcony as she planned. As with the climber case, we can pick out the causal connection between her decision for self-control and the movement of her body which causes the cake to drop over the balcony. However, despite this causal connection, our diabetic does not exercise self-control.

In both cases, we have a causal relation between thought and bodily movement. In the first example, the causal relation holds between the agent’s akratic thought and his driving the car. In the second case, it holds between the self-controlled thought and her movement of
dropping the cake off the balcony. According to the causal theorist, this relation should be enough to get us uncontrolled and self-controlled action. With the first, we have a break between the agent’s better judgment and his akratic thought which subsequently causes the action. With the second, we have an instance where the agent tries to overcome the break and exercise self-control and where her attempt succeeds in causing the relevant movement. But in both cases, the causal relation between the thought and the movement is not the right sort of relation. Thus, despite being caused by the antecedent psychic states of the agent, the movement still counts as mere movement, and not intentional action.\(^{30}\)

Both examples demonstrate the difficult position this model finds itself in by rigidly holding onto the causal picture of action. The defender of the standard approach I imagined tried to reject the first horn of the dilemma, insisting that the constitutive character of the causal relation alone can keep intentional action in view. According to this position, action is just bodily movement that has a particular kind of causal history in thought. And so in order to understand akratic action, this person argues, we also need to locate akratic failure in thought. However, the climber-style counterexamples block this response. Rather than make action an essential part of the account, the causal story in fact renders action completely problematic.

*Rational and Irrational Actions*

It would seem that the dilemma confronting the model of the break still stands. If the adherent of this approach sticks to her characterization of akrasia as a mental failure, then the role of action simply idles. If she rejects the second specification, then the action drops out altogether because we no longer have intentional action at all. On either horn, we seem to lose
our grip on the very phenomenon that the account hoped to explain in the first place, namely the possibility of akratic action.

    Rather than take either option, I propose that we give up this model altogether and so abandon the standard account. In urging a retreat from the model of the break, I want to suggest a new problem of akratic action: not the problem of responding to the negative thesis of Socrates and Hare, but rather how to account for the action in such irrational failure. Offering a satisfactory response to this problem is not so simple, particularly given our philosophical proclivities to look for explanations “in the head”. However, I believe we can take a step in the right direction when we notice that the model of the break gets its grip only if we already have the Davidsonian causal picture of action.

    To understand why, we can turn to Harry Frankfurt. Frankfurt argues that causal theories of action do not satisfy because they locate the difference between action and mere behavior simply in the causal histories of their mental antecedents and so they fail to take account of what a person does. According to such causal views, “actions and mere happenings do not differ essentially in themselves at all”—that is, they have no intrinsic differences—and so, he writes, “it is integral to the causal approach to regard actions and mere happenings as being differentiated by nothing that exists or that is going on at the time those events occur”. Rather, he claims, the difference between actions and other mere happenings lies at an earlier time among entirely different events.

    Like the causal theorist criticized by Frankfurt, the adherent of the standard approach finds no intrinsic difference between rational and irrational action and so locates the entire explanation of their difference extrinsically, in the causal etiology of what happens. Rather than
looking at what a person does at the time of such failure, those who defend the model of the break locate the difference at an earlier time, in the mental antecedents of the action. It is because the standard approach finds no intrinsic difference that Davidson can casually give an example of someone akratically brushing his teeth. The difference on Davidson’s account between my akratically brushing my teeth and my continently brushing my teeth lies not in anything to do with what happens when I brush my teeth, but in the mental break prior to the brushing. But it is precisely this assumption that the only relevant difference could be found in the causal antecedents that leads to the difficulties with the break in the first place. If we give up this assumption, then we can begin to resolve the new problem of akratic action by looking for the intrinsic difference(s) between rational and irrational action. 

To conclude, I am suggesting a new problem of akratic action—to figure out a way to take account of what a person does when she acts akratically. As I understand it, the standard approach makes no intrinsic distinction between rational and irrational actions. Rather, when an agent fails, this view locates the difference solely in her prior mental states. But by locating a break in her prior mental states, we lose sight of what the person does. In order to resolve this problem, we need to give up the model of the break. And in order to give up the model of the break, I maintain, we need to give up the causal account and the assumption that rational and irrational actions have no intrinsic difference. The problem of akratic action for ethicists and action theorists should then become the problem of explaining this difference.
1 I am grateful for comments on an earlier version of this paper from Agnes Callard, Anton Ford, Rafeeq Hasan, Erica Holberg, Gabriel Lear, Martha Nussbaum, Candace Vogler, and to audience members from a conference at Stanford University.

2 Although it is customary to translate *akrasia* as either weakness of will or incontinence, I believe both translations have unfortunate connotations. The former suggests some independent, executive faculty of the will while the latter makes the failure sound like a bladder control problem and so beyond the exercise of agency. In order to avoid these implications, I will continue to use the Greek term, although at times I will use the ‘failure of self-governance’ and the ‘failure of self-control’ interchangeably with the Greek.

3 Alfred R. Mele, *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control* (Oxford: Oxford University Press, 1982), 4-5 [original emphasis]. Mele is not alone with this thought. This distinction between akrasia as character vs. akrasia as action is a common move in the contemporary literature with most philosophers taking themselves to be talking about action. So, for example, in the introduction to a recent volume of collected papers on akrasia, Sarah Stroud and Christine Tappolet write that “Philosophical discussions have focused not so much on weakness of will as a character trait than on the sort of action that manifests it: roughly, intentional action contrary to one’s better judgment, that is, contrary to the judgment that another course of action would be better.” *Weakness of Will and Practical Irrationality*, ed. Sarah Stroud and Christine Tappolet (Oxford: Oxford University Press, 2003), 2.


5 Rorty, 333.

6 David Pears, *Motivated Irrationality* (South Bend, Indiana: St. Augustine’s Press, 1998), 125.

7 Pears, 107.

8 Pears, 14.

9 Pears, 6.


12 It is important to distinguish here “any model of the break on the standard account” from other versions of the break that fall outside the scope of my argument. The standard account, as I conceive it, holds the causal view of action that I expound in the next section, and so locates the break by identifying some occurrent thought which precedes the action and causes it to come about. However, in giving up the causal view, another identification of thought becomes available and so, too, does another model of the break. On this alternative conception, the relevant thought is not some occurrent thought that standards prior to the action and so causes it to come about. Rather, the relevant thought gets identified by appeal to a broader context over the course of an agent’s life—some larger, rational structure that underwrites the agent’s thought and actions over time. So for example, G. E. M. Anscombe identifies a break between what the agent does and his final ends. See “Thought and Action in Aristotle,” in *From Parmenides to Wittgenstein: Collected Philosophical Papers Volume I* (Oxford: Basil Blackwell, 1981), 66-77. For a similar line of thought, Philippa Foot and Sarah Broadie suggest a kind of break between the akratic agent’s action and his character. See Philippa Foot, *Natural Goodness* (Oxford: Oxford University Press, 2001); and Sarah Broadie,

13 Note to Wittgenstein Workshop: This second stage of the argument is not included in this version of the paper.

14 To be clear, it is actually important that the.akratic agent does not suffer from an explicit contradiction. Like theoretical forms of irrationality such as self-deception, most philosophers reject the idea that an agent could intelligibly suffer form an outright contradiction in beliefs, attitudes, and judgments.

15 The account that follows is drawn, more or less, from Davidson’s essays, “Agency” and “Actions, Reasons, and Causes,” in Essays on Actions and Events 2d ed. (Oxford: Oxford University Press, 2001).

16 In fact, Davidson thinks that all we ever do is move our body and that the rest “is up to nature.” See “Agency,” 43-62

17 “Philosophy as Psychology” in Essays on Actions and Events, 233.

18 I want to emphasize that on the standard account, actions are identical with the movement of a human body. Or to repeat the canonical example from contemporary philosophy of action, my raising of my arm is identical with the movement of my arm going up. But although such identification goes together with the standard account, this is by no means universally accepted. Jennifer Hornsby has argued that my action is distinct from the movement of my body. Rather, on her view, my action is an event which causes my body to move. “Agency and Actions,” in Agency and Action, ed. by John Hyman and Helen Steward (Cambridge: Cambridge University Press, 2004), 1-23. Likewise, Maria Alvarez and John Hyman argue both that my action is distinct from my body’s movement, and furthermore, that actions are not even events. “Agent’s and their Actions,” Philosophy 73 (1998): 219-245. These unorthodox conceptions of action present an interesting alternative to the standard account and so might create the potential for a different conception of the break. However, such possibilities fall outside the scope of my argument in this paper.

19 There are, of course, different things one might mean by a “concluding better judgment” and different ways to see one’s better judgment as a “conclusion.” This will become clear in the next section of the paper.

20 I am intentionally side-stepping various interpretive issues surrounding Socrates’ account. First, in some of the relevant passages Socrates includes the hypothetical: “Then if the pleasant is the good…” (emphasis mine). Many have thought that Socrates conclusion here thus depends on the commitment to the identity of pleasure and the good, i.e. “psychological hedonism,” which he attributes to “the many”. If that reading is correct, then Socrates’ argument would only entail a rejection of akrasia if one also holds the hedonistic hypothesis. This is the reading suggested by Gerasimos Santas in “Plato’s Protagoras and Explanations of Weakness,” The Philosophical Review 75 (1966): 3-33. Similarly, G. Klosko argues in “On the Analysis of Protagoras 351B-360E,” Phoenix 34 (1980): 307-322, that Socrates’ argument depends on a view about human motivation that he calls “egoism”. More recently, however, Michael Morris has claimed in “Akrasia in the Protagoras and the Republic,” Phronesis (2006): 195-229, that Socrates’ argument can be separated from this dubious hypothesis. However we read these problematic passages, it seems clear that Socrates does hold the view that “if someone were to know what is good and bad, then he would not be forced by anything to act otherwise than knowledge dictates” (352c) and so for the sake of simplicity I will refer to the Socratic position as a rejection of akrasia without the additional hedonistic or egoistic complication.

Second, I am ignoring the specific role that “knowledge” may play in Socrates’ view. The discussion of akrasia in the Protagoras appears in the context of a discussion about the unity of the virtues and seems to begin with the question whether or not knowledge is “a fine thing capable of ruling a person” or whether it can be “utterly dragged around…as if it were a slave” (352b-c). However, in the conclusion of his argument at 358b-c1 Socrates restates the claim in terms of knowledge and belief. Whatever significance knowledge may or may not play for the purpose of Socrates’ argument, contemporary discussions of akrasia that begin by invoking the Protagoras typically avoid the role of knowledge. Indeed, Donald Davidson takes it as an advantage of his view that he frees the problem of akratic action from any dependence on knowledge. See Davidson, “How is Weakness of the Will Possible?” in Essays on
Actions and Events, 21-42. Thus, for the sake of relevance in presenting the standard contemporary approach, I will refer to the Socratic thesis primarily in terms of belief or judgment.


22 On the standard view, you will recall, some bits of behavior count as an exercise of agency if it is caused by an antecedent mental state(s) which also rationalize the action. Davidson, for example, says that an action is caused by a belief-desire pair. See “Agency”. Michael Bratman, basically agreeing with Davidson, thinks that we also need to make room for intention as an additional mental state we have when we act intentionally. “Two Faces of Intention,” The Philosophical Review 93 (1984): 375-405. Some, of course, have thought that we cannot reduce the exercise of agency to causation by beliefs, desires, and intentions. For example, J. David Velleman has argued that these various accounts lose sight of the agent altogether. But even in trying to fit the agent back into the picture, Velleman proposes identifying the work of the agent with yet another antecedent mental state, a desire to act in accordance with reasons. See Velleman, “What Happens When Someone Acts?,” in The Possibility of Practical Reason (Oxford: Oxford University Press, 2000), 123-143. Thus, we can even include Velleman in the standard account insofar as any action that results from an exercise of agency will be caused by some antecedent mental state. Of course, akratic action on his account will turn out to be a defective kind of action, since it will not arise from an agent’s desire to act in accordance with reasons. In such cases, Velleman’s akratic acts more like the Davidsonian agent, with only antecedent states like belief and desire (and perhaps intention) involved.

23 Of course, even these could be described as resulting from prior mental states, only not ones in virtue of which they would count as full-blooded actions, or not even actions at all (in the case of startle responses). But if even these extremely defective pieces of behavior stand in such relations, then the case becomes even stronger that an akratic action cannot happen without a connection to some prior mental state(s).

24 Davidson, “How is Weakness of the Will Possible?”


26 Davidson also identifies our concluding evaluative judgment with intention. See his later essay, “Intending,” in Essays on Actions and Events, 83-102. Thus, Pear’s point against Bratman could just as easily be made against Davidson.

27 Specifically, Pears argues that the inner intention has the form, “I intend myself to φ” and the outer intention takes the form, “I intend to φ”. On his view, the akratic break arises between these two intentions because he thinks that we can intend ourselves to φ without thereby intending to φ, just as I can intend for Smith himself to φ without thereby causing Smith to intend to φ. See Motivated Irrationality.


29 “Freedom to Act,” in Essays on Actions and Events, 79.

30 Nor will it suffice to simply stipulate that the causal relation must be “of the right sort” as Davidson and other defenders of this approach have done. Responding to Davidson’s view, Anscombe writes: “He speaks of the possibility of ‘wrong’ or ‘freak’ causal connexions. I say that any recognizable causal connexion would be ‘wrong’, and that he can do no more than postulate a ‘right’ causal connexion in the happy security that none such can be found. If a causal connexion were found we could always still ask: ‘But was the act done for the sake of the end and in view of the thing believed?’” “Practical Inference” in Human Life, Action and Ethics: Essays by G. E. M. Anscombe, ed. Mary Geach and Luke Gormally (Exeter: Imprint Academic, 2005), 110-111.

31 Harry G. Frankfurt, “The Problem of Action,” in The Importance of What We Care About (Cambridge: Cambridge University Press, 1988), 69-70. Frankfurt offers one way in to understand the mistake that the standard
approach makes. However, his solution to the prior causal views does not really escape the problem he discovers. I address Frankfurt’s proposed solution in a longer version of this paper.

Anscombe offers us a further—and I think superior—way to understand the problem with the causal view. If we modify what she says against the causalist like Davidson, we could say that the problem arises from the “failure of percipience [in] the standard approach” by which we take action as “already given” and then specify that we are talking about irrational action rather than the rational. “Practical Inference,” 111. When Anscombe charges causal theories with treating action as a given, she urges us to see that action is not a mere byproduct of thought, a structureless, atomic entity that simply follows our judgment, whether the action be pulling the trigger of a gun, going for a walk, or building of a house. Rather, on Anscombe’s view, action itself has a rational, articulable structure, what she calls the A-D order of action [Intention (Cambridge: Harvard University Press, 1963), §25].

Michael Thompson has further elaborated Anscombe’s claim in his seminal essay, “Naïve Action Theory” [now published in Life and Action: Elementary Structures of Practice and Practical Thought (Cambridge: Harvard University Press, 2008), 85-146] by arguing that the action or event form needs to be understood as an unfolding process with the means-end or part-whole form.

In arguing that the model of the break ignores action, I mean to follow both Anscombe and Thompson with the idea that action—and by extension akratic action—must be understood, not by causal antecedents, but rather by the structure of the action itself.

32 Note to Wittgenstein Workshop: As this is a work in progress, I brought the paper to close at this point. However, the actual workshop version may differ.