

LIFE INTERRUPTED: AKRASIA, ACTION, & ACTIVE IRRATIONALITY

DISSERTATION SUMMARY
CHARLES NORMAN TODD

We understand all too well that sensible adult human beings lose self-control from time to time. For example, consider Edith who confronts a large, mouth-watering chocolate cake after rigidly maintaining her diet for weeks. As she longingly stares into the velvety icing, she finds herself drawn to it. Although she tries to resist knowing that it is the worst thing to do to her diet, she nonetheless caves in, devouring not just a slice, but the entire thing. Or imagine Rolf who becomes seriously irritated by a colleague at a staff meeting. He knows that confrontation in the past has produced uncomfortable tension and that his superiors will not look favorably at animosity in the office. Rolf knows he should just quickly leave the meeting and head back to his desk, but he nonetheless stops the colleague and immediately tries to put him in his place, leading to an uncomfortable and regrettable dispute. Both Edith and Rolf lose self-control and they do so knowing full well that they ought to act otherwise.

Philosophers have found this behavior fascinating and so have attempted to give it more precise definition. The ancient Greek philosophers called it *akrasia*, which literally means a lack of command over oneself. Others call it ‘incontinence’ or ‘weakness of will’, trying to signal that the agent has lost the full governance of her own action.¹

¹ I will generally use the Greek *akrasia* to refer to the phenomenon of losing self-control, and speak of the akratic agent and the akratic action respectively to refer to the doer and to what is done.

Whatever we call it, the philosophical consensus holds that the akratic agent knows or judges what she ought to do and yet purposively does otherwise. Although the agent loses the governance of her own action, the consensus also holds that the akratic agent still *acts intentionally*. Or to put it another way, this kind of behavior counts as intentional action even though it is not fully rational. So Edith breaks her diet, not just knowingly, but also intentionally. And Rolf gets into his office spat, not just against his better judgment, but as something that he intends to do. This emphasis on genuine action then allows us to distinguish the akratic agent from one that suffers from compulsion or that does not act freely. Edith and Rolf both have a choice to do what they know they ought to do, and yet they don't take it.

So we can summarize the phenomenon of akrasia as follows.

1. In order to count as akratic, an agent must act against her better judgment.
2. She must contravene that better judgment knowingly and not accidentally or in any other way that could be without her knowledge.
3. Her action must be done intentionally and so it must satisfy at least the majority of the criterion necessary for intentional action.
4. It must be done freely, and not under any kind of compulsion.

■

As I said, we understand this kind of irrational failure all too well given that it seems to plague our practical lives. But despite its commonplace, losing self-control raises some serious philosophical conundrums. For starters, it seems that if we really *know* what we ought to do in some situation then we will just do it. Some have thought that having knowledge just is having a special kind of power, one that differs from merely *supposing* what one ought to, or *hypothesizing* what one ought to do, or believing

what one ought to do. Socrates put forward this conception of knowledge as a special power in an agent in Plato's dialogue the *Protagoras*. Socrates says,

...knowledge is a fine thing capable of ruling a person, and if someone were to know what is good and bad, then he would not be forced by anything to act otherwise than knowledge dictates...²

In other words, knowledge is of the highest order, and so if you know the good thing to do—what you ought to do—then you really will do it. For to do otherwise is to suggest that you don't really know it. It appears then that Socrates denies the possibility of *akrasia*, of truly acting against what you know you ought to do.³

Socrates' puzzle focuses on the element of knowledge. But other philosophers explain the problem in terms of the connection between our better judgment and our motivation. For it seems to some that practical judgment implies a certain kind of motivation, i.e. that judgment and motivation are intrinsically linked.⁴ For example, consider a conversation between Hugh and Jan.

Hugh: "It seems like you have two options, A or B. What should you do?"

Jan: "I absolutely must do A, that is what I ought to do."

Hugh: "So when are you going to start A-ing?"

Jan: "Oh, I am not going to do A. I am going to B instead."

Hugh: "But you just said that you ought to do A!"

² Plato, *Protagoras*, 352c4-6.

³ The question of Socrates' actual view in the *Protagoras* is a complicated issue, one that I cannot address here. Because of the wider context of the dialogue, scholars have disagreed as to whether or not Socrates is giving his own view or merely engaging with the views of his interlocutors. For a influential discussion of this text, see Gerasimos Santas, "Plato's *Protagoras* and Explanations of Weakness."

⁴ Numerous philosophers take up this connection between judgment and motivation. For example, Michael Smith discusses it in *The Moral Problem*, specifically with respect to the question of *moral judgment*. And Sergio Tenenbaum goes to great lengths to argue that there can be no gap between our positive evaluative judgments of some action and our motivation to do that action in *Appearances of the Good*.

Jan: “I know I ought to do A, but I have no reason to do A and so I don’t want to.”

Now many philosophers would reject the possibility of an agent like Jan. They would insist that if she really judges that she ought to do A, then she will take herself to have a reason to do A. And if she takes herself to have a reason to do A, then she will in fact be motivated to do A—and not just motivated, but sufficiently motivated to do that action. The idea that we judge some action better than its alternative just seems to imply that we also want to do that action more than its alternative. But if judging or knowing what we ought to do implies motivational efficacy, then it would not be possible to ever act against our better judgment. This problem suggests that when Edith breaks her diet, it must be because she really judges that breaking her diet is the best thing to do.

Still other philosophers focus, not on conceptual connections between elements of our moral psychology, but on the force of the desires or pleasures that lead us to act akratically. It seems to these philosophers that if a desire can overcome an agent’s best rational efforts at self-control, then that agent must not be truly free. For example, David Pugmire writes,

[As] everything stood the desire does seem to have been as good as irresistible by him then...[It] did defeat his best efforts: he put himself through a deliberation that opened him, as much as anything in his power could, to what he was doing, and he reached a dissuasive all-things-considered value-judgment, resolved and set himself against what he then did anyway...What more could he have done? However, if the available resources for resistance failed, it would be arbitrary to insist that the desire was resistible...⁵

Pugmire’s point is that if the agent has already reached his better judgment as the result of rational deliberation, then the agent has already exhausted all his resources for

⁵ David Pugmire, “Motivated Irrationality II,” 188-189.

following through on that judgment. He cannot turn back to further rational deliberation, for by hypothesis the judgment already reflects the agent's best considered judgment on the matter. So if after all that the agent acts otherwise, then it must be because the desire was irresistible. And if it was irresistible, then the agent wasn't free.⁶

Finally, another form of the puzzle centers around the idea of the *causes* of intentional action. According to standard philosophical wisdom, all intentional actions are caused by a reason. When an agent acts, some belief and desire of the agent rationalize the action and cause it to come about. So if akratic actions are indeed intentional actions, then even akratic actions must be caused by such reasons, i.e. by a belief-desire pair which rationalizes the action. But what happens when we have reasons for different actions? Which one wins out? Davidson writes,

Causal theories of action are challenged by intentional actions that are contrary to the actor's best judgment. For if reasons are causes, it is natural to suppose that the strongest reasons are the strongest causes.⁷

So if, as he says, the strongest reasons are the strongest causes, then akratic action should not be possible. For whichever reason the agent ultimately acts upon, that one will have been the strongest reason.

Each of these puzzles recognizes some kind of problem that arises for akratic action, and they all take a slightly different approach.⁸ What they have in common is a shared sense of tension between intentional action, on the one hand, and its irrationality on the other. To count as irrational, the akratic agent must know or judge sincerely what

⁶ Alfred Mele has spent considerable time arguing against Pugnire and those who share his view that the akratic agent is in fact free to do otherwise and so that the agent's resources for resistance have not been completely exhausted. See Mele's *Irrationality*.

⁷ Donald Davidson, *Essays on Actions and Events*, xvi.

⁸ For a helpful discussion and summary of the various puzzles and criticisms made against them, see Arthur Walker, "The Problem of Weakness of Will."

she ought to do. But that knowledge or judgment seems to carry a special kind of weight, whether it is the power of knowledge or the motivational efficacy of judgment. Since the agent deploys her greatest exercise of reason in making that judgment, then it seems as if she must not be free should she do otherwise. And since stronger reasons seem like stronger causes, then surely the agent just does what she really thinks best. So each of these puzzles reaches the same conclusion: akratic action is not possible.

■

In order to avoid these puzzles, some philosophers think we need to distinguish between two different sorts of akrasia, what we might call ‘weak’ and ‘strong’ akratic action. In the weak case, the agent intentionally acts against her better judgment, but she does so because she doesn’t see things quite so clearly at the moment she acts. Perhaps she forgets what she ought to do, or perhaps she doesn’t understand the situation, or perhaps emotion or desire clouds her judgment. In Edith’s case we could imagine that hunger from dieting makes her temporarily forget her better judgment, or Rolf’s irritation might confuse his decision-making ability after the meeting. In such cases, we would still say that the agent acts against her better judgment in some sense, but not with full knowledge of what she does. By contrast, in the strong case, we say that the agent is fully aware that she acts against her better judgment.⁹ This agent stares what she ought to do in the face, so to speak, knowing she ought to do something else. In marking this distinction, some philosophers insist that we can avoid the puzzles because they only get traction with strong akrasia. But even if strong akrasia is not possible, we can still make sense of the weaker variety and that is all we really need.

⁹ This sort of akrasia has been given different names by different philosophers. Kieran Setiya calls it “clear-eyed akrasia” and Alfred Mele calls it “strict akrasia”. See Setiya, *Reasons Without Rationalism* and Mele *Irrationality*.

This idea of weak akrasia particularly gets its grip with some philosophers who work on virtue.¹⁰ For example, John McDowell has argued that having a virtuous character is akin to having a carefully attuned perceptual capacity. The virtuous man sees the world a certain way and so will be motivated accordingly, as when he sees a friend in need and so automatically sees a situation in which kindness is required. If strong akrasia is possible, then the akratic agent would see the situation in the exact same way as the virtuous man sees it, and yet not be motivated to act.¹¹ But McDowell says that the possibility of such strong akrasia undermines the very idea of virtue. The akratic agent *cannot* see the situation the same precisely because she is not virtuous. She doesn't see the situation as one in which kindness is required or else she would be motivated to act kindly. McDowell does think that people act akratically, but only because they do not see the situation in quite the same way as the virtuous man. Rather, the akratic agent suffers from a kind of blurry vision. Perhaps the salient features of the situation escape her. McDowell's akratic agent thus only suffers from weak akrasia

So we might accept one of the various puzzles and deny the possibility of akrasia, or we might allow the weak case and suggest that the agent becomes confused in some way or doesn't see the situation clearly. And yet, these two options aside, it simply seems an empirical matter that people very often judge what they ought to do while knowingly acting otherwise. Strong akrasia is a completely common and everyday phenomenon. People lose control all the time in countless ways and no amount of philosophical theorizing can dissuade us of that impression. In fact, David Wiggins goes so far as to say

¹⁰ John McDowell, "Virtue and Reason".

¹¹ John McDowell, "Are Moral Requirements Hypothetical Imperatives?", 80-81.

that “anyone not under the influence of theory” will recognize the phenomenon of akrasia and its characterization, and if your “pet theory” does rule it out, then so much the worse for your theory.¹² Thus, in order to do justice to our experience, we need an account of strong akrasia that makes it plausible and that doesn’t just water it down to some lesser failure.



It has become standard among philosophers who want to explain strong akrasia and so to avoid the puzzle to turn to the agent’s antecedent mental relations. The thought is that human behavior, both rational and irrational, can be explained by what an agent believes and desires before she acts. When all goes well, the standard approach says, an agent’s better judgment governs her action. But when she acts akratically, the better judgment and the action come apart, and some other attitude or judgment governs the action instead. According to the standard approach, the action counts as akratic in virtue of the conflict between the better judgment and that other mental state. Some describe that conflict as a kind of mental incoherence or inconsistency while others characterize it is a break or gap. Either way, the source and explanation of the irrationality rests with that mental relation. The goal for these philosophers is to present the right kind of mental conflict so that they avoid the paradoxical appearance of the behavior.

Now, depending on their preferred picture of moral psychology, most philosophers can come up with an account of that mental breakdown which avoids the puzzle. But whatever the picture, the various solutions will share a basic common assumption: namely, that the solution will be found by looking at a *static* mental relation.

¹² David Wiggins, “Weakness of Will Commensurability, and the Objects of Deliberation and Desire,” 241-241.

For example, Edith judges that she ought to stick to her diet and so she should avoid eating the cake. But she also has a desire to eat the cake. When she loses self-control and gives in to her desire, the philosopher will say she is akratic in virtue of the inconsistency between her better judgment and that desire. And *inconsistency* is a static relation. It characterizes a relation that obtains among some set of mental states—in this case, between a judgment and a desire. Such relations can of course come and go as when the offending desire wanes in strength or when the agent changes her mind about what to do. But while the relation obtains, its presence is something static and changeless.

This focus on static mental relations to explain akrasia has ramifications that might make us worry about the assumption. For starters, since the mental relation is *antecedent* to the action, it appears as if an agent has already lost self-control before she ever starts to act. Second, since the relation obtains *in full* before the action is complete, it appears as if the agent can be charged with complete irrationality, even if she later course corrects and regains self-control. And third, since the irrationality manifests as a *mental relation*, it seems as if the loss of self-control is on a par with other mental irregularities, such as bad reasoning, miscalculation, or inconsistent belief.

But more importantly, this assumption that philosophical explanations of akrasia reside among static mental relations neglects the undeniable fact that almost all akratic action, like all intentional action, takes time. Akratic actions do not simply emerge fully formed from the agent's head nor do they happen in an instant. Rather, they almost always becomes manifest through time as the action unfolds and as the agent struggles with self-control. Someone might start down a path towards akrasia, then start to regain self-control, and then lose it again, all within the same course of action. Or, an agent's

akrasia might only manifest slowly over the whole, without any one thing along the way counting as akratic on its own. The fact that akrasia manifests itself over time as an action unfolds sits rather uncomfortably with the idea that it is fully explained by a static mental relation which obtains before the agent ever starts to act.

To consider the previous examples, someone like Edith does not just manifest akrasia the second she forms the desire to eat the cake, or even the moment she lifts the fork. Rather, her failure of self-governance emerges over the course of her action. For example, she might take a first bite, but then put the cake back and try to reaffirm her commitment to her diet. She might go back to some other task only to return to the kitchen later for another bite. Before long, she finds herself plunging hard and fast into the cake. In the middle of this struggle she might go to great lengths to rid herself of the temptation. She might put the cake back in the fridge, she might wrap it tightly in foil and cover it with other items, or she might even brush her teeth so as to get the taste out of her mouth. Each of these attempts comes from her struggle with self-control. Of course, she might still give in and undo some of her own preventative measures. Ultimately, it is in the unfolding and changing action that her akrasia emerges and not simply in some static mental inconsistency that obtains prior to her acting.

We could imagine the same sort of unfolding irrationality with Rolf. He might start up towards his colleague when the meeting is over but then sit back down to wait a moment and let the room clear. But then as he gets up he sees the colleague still nearby and so he heads directly over to him. He knows he is in dangerous territory and close to creating a bad situation. So he tries to start off slowly and carefully, trying to stay calm, but the longer he talks the tenser they both become. He tries to back out of the

conversation and lower the growing animosity in the air. But then, at the last moment when he is about ready to end things, he ratchets the conversation back up, letting his irritation show through, and ends any chance of a peaceful outcome.

The examples of Edith and Rolf show us that not all akratic actions happen in a moment, nor can a single, static relation easily explain them. These agents are not just rational or irrational within the single, unified action. Rather, their rationality waxes and wanes throughout the action as the battle for self-control continues.

The negative aim of this dissertation is to argue that by neglecting the temporal dimension the standard approach cannot explain the majority of cases in which an agent struggles with and ultimately loses self-control. The positive aim is to develop a genuine alternative.

■

I begin my argument in Chapter 2 where I distinguish between active and static irrationality. I draw this distinction by deploying the metaphysical distinction between states and events. A state doesn't change over time, but rather remains the same. A state doesn't have parts or phases. And any interval of time from a state will be identical with the entire state. By contrast, an event exhibits change over time and so doesn't remain the same. An event has parts and phases, e.g. it typically has a beginning, middle, and end. And any interval of time from within a state will be different from the whole.

Using this state/event distinction, I claim that static irrationality characterizes states, whereas active irrationality characterizes events. Static irrationality either obtains or fails to obtain at any particular moment, whereas active irrationality unfolds through time. And static irrationality characterizes an action homogeneously throughout the entire action, whereas active irrationality allows that the action can change with respect to its

irrationality through time. Since akratic action, in the main, typically manifests in an extended struggle the agent has with self-control, I show that while the standard approach treats akrasia as a static, mental moment, we must instead locate akrasia in the active variety.

Opposed to the dynamic and changing character of akratic action is the view I call *incoherentism*, the philosophical view that practical and theoretical irrationality arises from incoherence or inconsistency among an agent's mental states. Donald Davidson is the foremost proponent of this view, claiming that irrationality is "the failure, within a single person, of coherence or consistency in the pattern of beliefs, attitudes, emotions, intentions and action."¹³ Such *patterns* among a set of attitudes are the paradigmatic case of static irrationality. Although the pattern will change as the attitudes change, the relation of coherence at any one time will be either rational or irrational. And it is that static relation which the incoherentist thinks explains akrasia. So Davidson writes that, "strictly speaking...we might want to say 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."¹⁴ Although the incoherentist approach has broad application to various kinds of error in our thinking, my claim is going to be that the incoherence of this view is only appropriate for characterizing the irrationality of states of agents, not changing events.

In Chapter 3 I present and criticize the standard approach to akrasia among contemporary philosophers. This view, what I call the *model of the break*, explains

¹³ Donald Davidson, "Paradoxes of Irrationality," 170.

¹⁴ Davidson, "Incoherence and Irrationality," 194.

akrasia by identifying a break or gap between two different mental states. Amelie Rorty first used the language of the “akratic break,” describing a picture in which thought progresses through time towards action.¹⁵ When all goes well, the agent’s better judgment will cause intention which will then cause the action. But when the agent acts akratically, there is a juncture in the stages of thought in which a break opens up, creating what she calls “psychological akrasia”. David Pears offers similar language, speaking of the agent as making a “jump” over a mental barrier, and Sebastian Gardener calls it a mental “propositional fracture”.¹⁶ But it was Davidson’s earlier and famous essay “How is Weakness of the Will Possible?” that turned searching for the break’s location into a philosophical pastime.¹⁷ Although he did not call it a break, he did try to identify the precise mental location where the agent became akratic, as if all it took was a simple mental error and then the agent was instantly irrational. Philosophers following in Davidson’s path agreed with his assumptions about the explanation of akrasia, but they disagreed about where precisely the break occurred.

Despite its popularity and apparent success, I argue that this model treats akrasia as a purely static, mental failure that happens before the agent acts and so completely ignores the temporally extended struggle. If all it takes for an agent to suffer from akrasia is a mental break that arises in one’s thought prior to acting, then the temporally unfolding action plays no role in an account of that irrational failure. And if the action plays no explanatory role, then the standard approach completely ignores the dynamic

¹⁵ Amelie Rorty, “Where does the Akratic Break Take Place?”

¹⁶ David Pears, *Motivated Irrationality*, 125; Sebastian Gardner, *Irrationality and the Philosophy of Psychoanalysis*, 35.

¹⁷ Davidson, “How is Weakness of the Will Possible?”

and on-going struggle an agent has with self-control. Even if some akratic actions may be explained by a momentary break, the majority of them cannot.

However, giving up the model of the break is not so easy, particularly when it comes packaged together with other common assumptions in the philosophy of action. In particular, I show that the model of the break is inextricably linked with Davidson's causal theory, which views action as an event brought about by a rationalizing antecedent mental cause. The causal theory bizarrely leads us to the conclusion that fully rational actions are fundamentally indistinguishable from akratic actions in which an agent loses self-control. Edith could rationally eat a chocolate cake or she could akratically eat that cake, but the only differentiating mark of rationality according to Davidson will come from the antecedent causes. But this conclusion is the exact opposite conclusion we would draw if we attended to akrasia as an extended struggle with self-control, rather than to prior static states of mental inconsistency. So in this chapter I argue not only that we should give up the model of the break, but also that we need to fundamentally alter how we approach the philosophical problem of akrasia.

I offer my alternative account in Chapter 4, beginning with the rational structure of all intentional action. I follow G.E.M Anscombe's contention that all intentional action has a temporally extended rational structure. Any intentional action can be described in various ways, as an A-ing, a B-ing, a C-ing, a D-ing, and so on. Anscombe argues that these various descriptions are not just different actions, but that they hang together as a whole. The descriptions stand to one another as means to ends, so the agent is A-ing in order to B, B-ing in order to C, and C-ing in order to D. So this A-D order, as Anscombe calls it, provides us with a rational structure of action. It is in light of the agent's ends that

the various means are rationalized. Furthermore, that structure provides us with the temporally unfolding means-end structure of the action.

According to my view, akrasia is a kind of irrational deformity of that temporally unfolding means-end structure. Using the A-D order, we can characterize all actions as wholes consisting of temporal parts done for the sake of the whole. When all goes well, an action manifests the agent's rationality because the parts conduce to that whole as means to ends. But in akratic action, I show that the means and ends come apart, either because the means disrupt the agent's end in some way or because the agent prevents herself from undertaking means to her end at all. This coming apart of means and ends is something that can only be articulated over time and so explains why most akratic actions belong to active irrationality.

Finally, I show how this temporal account offers a much richer and more nuanced view of akrasia. Once we attend to the unfolding structure of action we will have a more complex picture of akratic failure. Akrasia does not come in a one-size-fits-all category that is identified by a single akratic break. Rather, depending on how the means and ends come apart, the kind of failure can actually look quite different even while still counting as akrasia. I thus offer a general schema for the varieties of akrasia. While this schema is not meant to be exhaustive, I hope that it allows us to appreciate just how messed up human action can be.