

Forthcoming in *Philosophical Psychology*

Multiplicity, Self-Narrative, and Akrasia

Jeff Sebo
New York University

Abstract

In this paper I present a new account of akrasia based on the idea that human psychology and self-narrativity are more complex and layered than we have traditionally thought. I begin by arguing that, if we have at least some different beliefs, desires, preferences, etc. in different situations, then we can rationally do what we think, at the time of action, is best for, or from the standpoint of, “part of me” while acting contrary to what we think, at the time of action, is best for, or from the more comprehensive standpoint of, “me.” I then argue that many of us do, in fact, think and act this way in everyday life, and that this kind of action satisfies all the criteria for akrasia. Finally, I briefly argue that, on my account of akrasia, akratic actions are not necessarily irrational or blameworthy, though they often will be.

1. Introduction

Many philosophers feel inclined to accept two claims that seem mutually incompatible. The first is that we sometimes act akratically, i.e., we sometimes act contrary to what we think, at the time of action, we have most reason to do. And the second claim is that every action is based on a “best judgment,” i.e., a judgment about what we have most reason to do. Most of the literature on akrasia and weakness of will, then, is about which of these claims we should reject, and why: Should we deny that akrasia is possible, or should we deny that every action is based on a best judgment?

Some philosophers take the first approach: they deny that akrasia is possible. For example, R.M. Hare (1952), Richard Holton (1999), Frank Jackson (1984), and Gary Watson (1977) argue that, when we *appear* to be acting akratically, we are actually doing something else instead. For instance, we might be acting contrary to what we think *other people* think, at the time of action, is best. We might be *compulsively behaving* contrary to what we think, at the time of action, is best. We might be acting contrary to what we thought, *before* the time of action, is best. And so on. This approach preserves the link between action and best judgment, but it fails to vindicate the phenomenology of at least some apparent cases of akrasia – cases where we really do seem to be *acting* (rather than merely behaving) contrary to what *we* (rather than merely other people) think, *at* the time of action (rather than merely *before* the time of action) is best.

Other philosophers, in contrast, take the second approach: they deny that every action is based on a best judgment. For example, Donald Davidson (1969/2001) and Michael Bratman (1979) argue that every action is based on a judgment about what we have *some* reason to do, but

not necessarily on a judgment about what we have *most* reason to do. This approach vindicates the phenomenology of these apparent cases of akrasia: it allows us to say that we sometimes act contrary to what we think, at the time of action, is best. But it also severs the link between action and best judgment, which complicates our theories of action and agency and arguably creates more problems for these theories than it solves.

My aim in this paper is to develop a new approach to the akrasia problem that tries to have it both ways. Specifically, I will argue that we can rationally make *two* best judgments at the time of action, one of which is more comprehensive than the other, and we can rationally act on the *less* comprehensive of these two best judgments. Thus, I will argue, we can vindicate the possibility of akrasia *whether or not* we sever the link between action and best judgment.

Specifically, we can say that we act akratically if we do what we think, at the time of action, we have most reason to do according to one best judgment, while acting *contrary* to what we think, at the time of action, we have most reason to do according to another, more comprehensive best judgment.

Of course, one might worry that this approach is a non-starter, for at least two reasons. First, one might worry, if we make two evaluative judgments at the time of action, and if one of these evaluative judgments is less comprehensive than the other, then the less comprehensive evaluative judgment is, by definition, not a *best* judgment at all but rather only a *prima facie* evaluative judgment. Second, one might worry that even if we *could* make two best judgments at the time of action, we would be rationally required to act on the *more* comprehensive of these two best judgments. But I think that this is too quick. The key is to see that we can rationally have two *self-conceptions* at the time of action, one of which is more comprehensive than the other, and we can rationally do what we think is best for, or from the standpoint of, the less

comprehensive of these two “selves”. For example, you can rationally do what you think, at the time of action, is best for, or from the standpoint of, “me as a person” while acting contrary to what you think, at the time of action, is best for, or from the more comprehensive standpoint of, “us as a family.” Similarly, I will argue, if you have at least some different beliefs, desires, habits, etc. in different contexts, then you can rationally do what you think, at the time of action, is best for, or from the standpoint of, “the part of me currently in control” while acting contrary to what you think, at the time of action, is best for, or from the more comprehensive standpoint of, “me as a person.”

I will proceed as follows. In section 2, I will argue that, if you have at least some different beliefs, desires, preferences, and so on in different contexts, then you can rationally (a) attribute mental states to “part of me” as well as to “me”, (b) make a different best judgment for, or from the standpoint of, “part of me” and “me” at the time of action, and (c) act on the former, less comprehensive best judgment rather than on the latter, more comprehensive best judgment. In section 3, I will argue that many of us do, in fact, think and act this way in everyday life, though we might not always realize it. In section 4, I will argue that this kind of action satisfies all the criteria for akrasia and explain how my account of akrasia relates to other prominent accounts. Finally, in section 5, I will briefly consider the normative implications of my account of akrasia. Specifically, I will argue that akratic actions are not necessarily irrational or blameworthy, though they often will be.

2. A Simple Case

I begin by arguing that, if you have at least some different beliefs, desires, preferences, etc. in different contexts, then you can rationally (a) attribute mental states to “part of me” as well as to “me”, (b) make a different best judgment for, or from the standpoint of, “part of me” and “me” at the time of action, and (c) act on the former, less comprehensive best judgment rather than on the latter, more comprehensive best judgment.

Here is a very simple version of the kind of case that I have in mind. Suppose that you have at least some different beliefs, desires, preferences, etc. at work and at home. For example, when you go to work, you tend to value your career a bit more than your family, you tend to think that you should prioritize your career over your family, and so on. And then, when you go home, you tend to value your family a bit more than your career, you tend to think that you should prioritize your family over your career, and so on. In this case we can usefully describe you as having a “work self” and a “family self” that stand in the same kind of relation to each other, and to the person as a whole, that, say, your *family members* stand in relation to each other, and to the family as a whole, i.e., they disagree about how to live (at least with respect to some issues), and yet they also have to live *together*, and therefore they have to compromise and coordinate with each other in order to get along.¹

Now, given that your work self and family self have to compromise and coordinate in order to get along, they might find it useful to accomplish this aim in the same kind of way that your family members might: by thinking about what kinds of shared practices represent a fair compromise between them, and then attributing a preference for these shared practices to themselves as a *group* (i.e., to “us as a family” in the interpersonal case and “me as a person” in the intrapersonal case). For example, your work self and family self might each find it useful to think, “Well, if my work self values work a bit more than family and my family self values

family a bit more than work, then *I as a person* value work and family in equal measure overall.” And then, having constructed a shared self-conception in this kind of way, your work self and family self might find it useful to compromise and coordinate in everyday life by doing what “I” prefer to do overall rather than what “the part of me currently in control” prefers to do all else equal. For example, suppose that your family self has to decide what time to leave for work in the morning. Even though they might prefer to leave at 10am all else equal, if they have a habit of thinking and acting from the standpoint of “me as a person,” then they might choose to leave at 9am instead, thereby acting in a way that maximizes your preference satisfaction across contexts without having to think about it in these terms at all.²

Granted, it might seem strange for your work self and family self to use the first-person singular to refer to what amounts to a collective self-conception. Why not instead do what we do in the interpersonal case: use the first-person singular to refer to their individual self-conceptions and the first-person plural to refer to their shared self-conception? But I think that this use of the first-person singular makes sense when we consider the fact that your work self and family self have to act on behalf of the person as a whole much more often than not, and therefore they have reason to think of the person as a whole as the “default” unit of practical concern for them. And in any event, I am not claiming here that your work self and family self are, in this simple case, rationally *required* to construct a shared self-conception or to use the first-person singular to identify with it. I am rather only claiming that they are rationally *permitted* to think and talk this way if they find it useful for purposes of compromising and coordinating with each other in everyday life.

Now consider the second part of my claim: if you attribute mental states to “part of me” and “me” in this kind of way, then it might sometimes be rational for you to make a different

best judgment for, or from the standpoint of, “part of me” and “me” at the time of action. This claim is easy to establish. If you have at least some different beliefs, desires, preferences, etc. in different contexts, and if you use the first-person singular to refer to an epistemic and practical standpoint that represents a fair compromise among them, then it might sometimes be rational for you to think the following kind of thought: “Even though *I as a person* have most reason to *x*, *the part of me currently in control* has most reason to *y*.” For example, consider the case where you have to decide what time to leave for work in the morning. If your family self is out at the time of action, and if your family self values staying in bed more than they value compromising and coordinating with your work self, then it might, in this particular case, be rational for your family self to think, “Even though *I* have most reason to leave for work at 9am, *the part of me currently in control* has most reason to leave for work at 10am.” (This is what it means for you to make a different best judgment *for* “part of me” and “me” at the time of action.)

Similarly, if you have at least some different beliefs, desires, preferences, etc. in different contexts, and if you use the first-person singular to refer to an epistemic and practical standpoint that represents a fair compromise among them, then it might sometimes be rational for you to think the following kind of thought as well: “Even though I as a person *think* that I have most reason to *x*, I *actually* have most reason to *y*,” and/or “Even though I as a person *think* that the part of me currently in control has most reason to *x*, the part of me currently in control *actually* has most reason to *y*.” For example, consider the case where you have to decide what time to leave for work in the morning. If your family self is out at the time of action, and if your family self rationally believes that your work self is not an epistemic peer with respect to this issue, then it might, in this particular case, be rational for your family self to think, “Even though I *think* that the part of me currently in control has most reason to leave for work at 9am, they *actually* have

most reason to leave for work at 10am.” (This is what it means for you to make a different best judgment *from the standpoint of* “part of me” and “me as a person.”)

Of course, it will not *always* be rational for you to make different best judgments for, or from the standpoint of, “part of me” and “me” in this kind of way. On the contrary: given that your work self and family self presumably (a) have many desires and preferences whose satisfaction requires them to compromise and coordinate with each other and (b) are epistemic peers with respect to many issues, they will generally speaking have very strong *prima facie* reason to make their individual best judgments conform to their shared, or collective, best judgment, once the latter has been established. (More on this in section 5.) Still, insofar as you attribute different mental states to the person as a whole than you attribute to your current self as *part* of this person, and insofar as you use the first-person singular to identify with the *former* self-conception rather than with the *latter* self-conception, it will at the very least *sometimes* be rational for you to make different best judgments in this kind of way.

Finally, consider the third part of my claim: if you make a different best judgment for, or from the standpoint of, “part of me” and “me” at the time of action in this kind of way, then you can rationally act on the former, less comprehensive best judgment while acting contrary to the latter, more comprehensive best judgment. This claim is easy to establish as well. In the simple case that we have been considering, your conception of “the part of me currently in control” describes who you are and why you do what you do at the time of action, whereas your conception of “me as a person” merely describes who you are and why you do what you do *in general*, or *on average*. Thus, insofar as you rationally make a different best judgment for, or from the standpoint of, “the part of me currently in control” and “me as a person,” you are rationally permitted, if not rationally required, to act on the former, less comprehensive best

judgment rather than on the latter, more comprehensive best judgment. The analogy with interpersonal groups is helpful here. For example, we all agree that it would be irrational for you to think “Even though *we* have most reason to move to California, *I* have most reason to try to get us to move to New York instead” and then buy a one-way ticket to California on the basis of this judgment. And in the same kind of way, it would also be irrational, in the simple case we have been considering, for you to think “Even though *I* have most reason to keep my current job, *the part of me currently in control* has most reason to take the promotion instead” and then sign a contract to keep your current job on the basis of this judgment. Likewise, we all agree that it would be irrational, all else equal, for you to think “Even though *we* think that I have most reason to move to California, *I actually* have most reason to move to New York” and then buy a one-way ticket to California on the basis of this judgment. And in the same kind of way, it would also be irrational, all else equal, in the simple case we have been considering, for you to think “Even though I *think* that the part of me currently in control has most reason to leave for work at 9am, they *actually* have most reason to leave for work at 10am” and then leave for work at 9am on the basis of this judgment.

The upshot is that if you have at least some different beliefs, desires, preferences, etc. in different contexts, then you can, in at least some cases, rationally use the first-person singular to refer to an epistemic and practical standpoint that represents a fair compromise among them; make a different best judgment for, or from the standpoint of, “part of me” and “me” at the time of action; and act on the former, less comprehensive best judgment rather than on the latter, more comprehensive best judgment.

3. Multiplicity and Self-Narrativity

I will now argue that many of us do, in fact, think and act this way in everyday life, though we might not always realize it. Of course, I will not be able to provide a full defense of this empirical claim here. But I will try to motivate it by showing that it plausibly follows from two other, more familiar empirical claims that many philosophers and psychologists already accept: that our personalities tend to be complex and dynamic, and yet we tend to *think* of our personalities as relatively simple and unified.

First, our personalities tend to be complex and dynamic, in the sense that many of us tend to have different personality traits in different situations.³ This can happen for a variety of reasons. For example, we play different social roles in different social contexts, and we develop different thoughts, feelings, and habits for each role. Also, our psychological dispositions are sensitive to seemingly innocuous features of the environment, e.g., how likely we are to help or hurt somebody might depend on whether we recently smelled fresh bread or found change in a pay phone. And of course, our psychological dispositions are also sensitive to how tired we are, how hungry we are, how drunk we are, and a thousand other variables. The general upshot is that we have many different, overlapping “selves” or “sides” or “parts” to our character that we cycle through many times each day, month, or year.

But even though our personalities are complex and dynamic in this kind of way, many of us tend to *think* of our personalities as relatively simple and unified, because we construct our self-narratives by thinking about who we are and why we do what we do *on average* or *in general*. Daniel Dennett (1992), Owen Flanagan (1996), and others have fleshed out this idea by saying that many of us construct our self-narratives with the aim of finding our “centers of narrative gravity,” and I will use the same metaphor here – though, for reasons that I explain

elsewhere, I prefer the term “center of psychological gravity” and will use that latter term instead.⁴ So, what is a center of psychological gravity? Think about it as the psychological equivalent of a center of gravity. You want to know what a particular agent (say, you) will do. How do you proceed? Well, given that the *actual* story about who you are and why you do what you do is incredibly complex, you find it useful to tell a simpler, “semi-fictional” story about yourself instead: specifically, you suppose that you have a single, unified set of personality traits around which all of your actual, fluctuating personality traits are evenly distributed. And then you try to explain, predict, and control (i.e. choose) your behavior by thinking about what an agent with this single, unified set of personality traits would do. Thus, for example, you might explain why you have a policy of working 8 hours a day (as opposed to, say, 7 or 9) by saying that you value work and family in equal measure. Of course, at some level you might realize that this story is too simple to be fully accurate, but for many purposes this simple, unified story is close enough to the truth and, if anything, more useful than the fully accurate story would be, since it provides you with a single, unified epistemic and practical perspective that you can identify with in order to explain, predict, and control your behavior in general.

So far everything that I have said should be familiar to people who follow the literature on personality psychology and self psychology: our personalities are complex and dynamic, and yet our self-narratives are (relatively) simple and unified. But here is where my story gets a bit more complicated. I think that, in addition to constructing simple, unified, “general” self-narratives that find our centers of psychological gravity *across* situations, many of us also, at least implicitly, construct complex, dynamic, “situational” self-narratives that find our centers of psychological gravity *within* particular situations. Why do we do this? Because we need different kinds of self-narrative for different epistemic and practical purposes. For example, as we have

seen, in some cases we need to be able to explain, predict, and control our behavior *in general*, e.g., you might need to explain why you have a general policy of working 8 hours a day. And in these cases it makes sense for us to explain, predict, and control our behavior in terms of a simple, unified, “general” self-narrative that finds our center of psychological gravity *across* situations. But in other cases we need to be able to explain, predict, and control our behavior *in particular situations*, e.g., you might need to explain why you missed dinner last night. And in these cases it makes sense for us to explain, predict, and control our behavior in terms of a complex, dynamic, “situational” self-narrative that finds our centers of psychological gravity *within* particular situations. Thus, for example, in addition to thinking of yourself as valuing work and family in equal measure overall, you might *also* think of yourself as having a work self who values work a bit more than family and as having a family self who values family a bit more than work. “Why do you usually work from 10-6?” “Because I value work and family in equal measure.” “Well then why did you get home so late last night?” “Well, *my work self* values work a bit more than family...”⁵

Now, notice that if you have different self-narratives at different levels like this, then each self-narrative will support a different interpretation of a given action: your general self-narrative (i.e., the self-narrative according to which you have a single, unified self across contexts) will support an interpretation according to which *the person as a whole* performed this action, and your situational self-narrative (i.e., the self-narrative according to which you have different, overlapping selves in different contexts) will support an interpretation according to which *your current self* performed it. And even though these general and situational self-narratives might not conflict *in principle* (since they are, at bottom, simply different stories that you tell yourself for different purposes), they might still conflict *in practice*. This can happen in

either of two ways: first, you might use one of these stories to explain a particular action when the other would be more appropriate. For example, you might use your general self-narrative to explain why you had a few extra drinks last night when your situational self-narrative would be more appropriate, with the result that you explain your action *not* by saying “My drinking self wanted to have a few extra drinks last night” but *rather* by saying either “I decided to have a few extra drinks last night even though I wanted to call it a night” or “Something made me have a few extra drinks even though I wanted to call it a night.” And not only are these latter explanations not very explanatory (especially the first of the two), but they are also positively *distorting*: the first accuses you of acting irrationally and the second accuses you of not acting at all.

The second way that your general and situational self-narratives might conflict in practice is that you might use both at once, in a way that makes your experience and interpretation of your behavior seem confusing, even paradoxical: for example, you might think (from the standpoint of your current self) that you have *most* reason to do what you think (from the standpoint of the person as a whole) you have only *some* reason to do. And therefore you might find yourself, say, choosing to have a few drinks even though you think (from the standpoint of the person as a whole) that you have most reason to call it a night instead.

These kinds of conflict are especially likely to arise in real-life cases, given two things. First, unlike in the simple case that we started with, most of us do not have, say, two distinct selves who are fully aware of the psychological differences between them. Instead, we have many overlapping selves who are, at best, partially aware of the psychological differences among them. Why are they only partially aware? There are a variety of genetic and social reasons for this. For example, we have a natural tendency to remember complex, dynamic phenomena as

having a simple, unified *average* of the properties and values that they actually have, and this natural tendency causes us to remember, and therefore think of, our own personality traits as simpler and more unified than they actually are.⁶ Moreover, we are socialized to think that we should have a coherent and stable set of beliefs, desires, preferences, etc., and this social pressure causes us to (charitably) think of our personality traits as simpler and more unified than they actually are as well.⁷ And the upshot of these genetic and social pressures is that we might sometimes misinterpret our general and situational self-narratives in everyday life: in particular, we might sometimes mistakenly think of our simple, unified general self-narratives as describing who we are and why we do what we do at the time of action (when in fact they merely describe who we are and why we do what we do in general, or on average); and we might sometimes mistakenly think of our complex, dynamic, situational self-narratives as describing a mere *part* of who we are and why we do what we do at the time of action (when in fact they describe all of who we are and why we do what we do at the time of action). Consequently, we might sometimes feel justified in using our general self-narrative to explain, predict, and control what we do in particular situations – a tendency which is usually harmless but which can also be misleading in the kinds of cases that we are considering here.

The second, related reason why these kinds of conflict are especially likely to arise in real life is that, unlike in the simple case that we started with, most of us do not clearly distinguish our general and situational self-narratives in everyday thought and language. For example, instead of always using the first-person singular to refer to the person as a whole and “part of me” to refer to our current self, many of us often use both terms in both ways, often in the very same sentence (e.g., “I hate the part of me that comes out around my parents. Can you ever forgive me for saying all those awful things?”) And to make matters even more complex, we

often use “part of me” to refer not only to our current self (as part of the person as a whole) but also to *part* of our current self (e.g., “What do I want to eat right now? Well, part of me feels like Chinese and part of me feels like Thai...”). The upshot is that, in at least some cases, our ways of thinking and talking about ourselves and each other are hopelessly multiply ambiguous, and this is likely to add to the confusion that comes from our constructing different, competing, stories about who we are and why we do what we do for different practical and epistemic purposes. For example, you might think, from the standpoint of your current self, that you have most reason to watch a movie right now, and then you might think, from the standpoint of the person as a whole, that you have most reason to grade papers right now. And insofar as you fail to distinguish these perspectives, you will have a strange kind of experience: the latter set of reasons will seem (and actually be) more *comprehensive* than the former set of reasons, and yet the former set of reasons will seem (and actually be) more *action-guiding* for you at the time of action than the latter set of reasons. And therefore you might find yourself doing what you think, at the time of action, is best for, or from the standpoint of, “part of me” while acting contrary to what you think, at the time of action, is best for, or from the standpoint of, “me” – while at the same time mistakenly interpreting “me” as the agent of this action, and therefore as acting irrationally or behaving compulsively.

Thus, when I say that many of us think and act this way in everyday life, I am not necessarily saying that we *think of ourselves* as thinking and acting this way in everyday life. Indeed, I think that very few of us have a fully accurate conception of our self-conceptions. But this is fine: my account of akrasia will not require us to have this kind of second-order self-awareness. In fact, as I will argue in a moment, I think that at least some akratic actions seem so confusing to us precisely because we *lack* this kind of second-order self-awareness: that is, in at

least some cases of akrasia, we experience and interpret ourselves as acting contrary to what we think, at the time of action, is best *precisely because* we fail to fully distinguish our general and situational self-narratives in our self-experience and -interpretation. And if this is right, then, once we clarify the structure of akratic actions for ourselves, akratic actions might start to seem less confusing to us than they currently do.

4. Akrasia

I now have everything that I need in order to present my account of akrasia. On my view, you act akratically if

1. You have at least some different beliefs, desires, habits, etc. in different situations;
2. You attribute mental states to *yourself as a person* by, e.g., finding your center of psychological gravity *across* contexts as well as to *your current self* by, e.g., finding your center of psychological gravity *within* this context;
3. You make a different best judgment for, or from the standpoint of, your current self and the person as a whole at the time of action;
4. You do what you think is best for, or from the standpoint of, your current self at the time of action while acting contrary to what you think is best for, or from the standpoint of, yourself as a person at the time of action; and
5. (optional) You interpret yourself as a person as the agent of this action, and you therefore interpret yourself as a person as either (a) irrationally acting contrary to what they think, at the time of action, is best, or (b) compulsively behaving contrary to what they think, at the time of

action, is best. (I say that this step is optional because, while it explains the phenomenology of at least some akratic actions, it is not, I think, necessary for an action to be akratic in the first place.)

I think that this account of akrasia satisfies all of the main criteria for akrasia – or at the very least, it satisfies them more than other accounts of akrasia and weakness of will do. First, it says that akrasia is an *action* rather than a mere *behavior*. Second, this account says that an akratic agent acts contrary to their most comprehensive evaluative judgment *at* the time of action rather than merely *before* the time of action. Third, this account says that akratic actions are based on best judgments, and therefore it vindicates the possibility of akratic action without severing the link between action and best judgment. Finally, and perhaps most importantly, this account explains the confusing, even paradoxical, phenomenology of at least some akratic actions, according to which the akratic agent experiences themselves as having *most* reason to do what they have only *some* reason to do. The idea here is that we sometimes experience and interpret the person as a whole as the agent of our actions (even though they are, in fact, a mere abstraction) and we sometimes experience and interpret our current self as a mere *part* of the agent of our actions (even though they are, in fact, the agent of our actions). And insofar as we do this, the phenomenology of akratic actions will be misleading: when we do what we think, at the time of action, is best for, or from the standpoint, of our current self while acting contrary to what we think, at the time of action, is best for, or from the standpoint of, the person as a whole, we will inaccurately experience and interpret the person as a whole as either (a) irrationally acting contrary to what they think, at the time of action, is best or (b) compulsively behaving contrary to what they think, at the time of action, is best.

Let me now go into a bit more detail about each of these points by showing how my account of akrasia compares to other well-known theories of akrasia and weakness of will. To be clear, my aim in what follows is not to argue that these other theories are false. Indeed, I think that we use the terms ‘akrasia’ and ‘weakness of will’ in many different ways, and therefore we need many different theories of akrasia and weakness of will to cover all of our uses of them. However, I also think that my theory of akrasia captures the kind of action that has proven so problematic for philosophy of action better than these other theories do, and I will argue for this somewhat more modest conclusion here.

First, Hare (1952) says that an agent is weak-willed if they act contrary to what they think *other people* think, at the time of action, is best – and explain themselves by saying that they are acting contrary to what *they* think, at the time of action, is best in an externalist, “inverted-commas” sense of ‘best’.⁸ For instance, suppose that everyone in your community but you thinks that drugs are bad. And suppose that you sometimes use normative concepts like ‘best’ to refer to what *the community as a whole* values instead of what you value. In this case, if and when you decide to, say, smoke a bit of pot, you might describe yourself as acting contrary to what you think, at the time of action, is “best” – where what you really mean is that you are doing what you think is best for, or from the standpoint of, yourself as a person while acting contrary to what you think is best for, or from the more comprehensive standpoint of, the community as a whole. My account of akrasia is very similar to this, except it says that the “community” whose best judgment you are acting contrary to is *yourself as a person*. So an analogous case would be: you decide to smoke a bit of pot and explain yourself by saying that you are acting contrary to what you think, at the time of action, is best – where what you really mean is that you are doing what you think, at the time of action, is best for, or from the standpoint of, your current self while

acting contrary to what you think, at the time of action, is best for, or from the more comprehensive standpoint of, yourself as a person. Structurally, then, these two explanations of apparent cases of akrasia are the same. But since we all identify with our “intrapersonal communities” much more than with any given interpersonal community, my account of akrasia explains why, if and when we act akratically, we (confusingly) experience ourselves as acting contrary to what we think, at the time of action, is best in a not fully internalist but also not fully externalist sense of ‘best’ – which is why I am calling my account of akrasia an *account* rather than merely an error theory.

Second, Watson (1977) says that an agent is weak-willed if they merely *behave* contrary to what they think, at the time of action, is best.⁹ For example, suppose that you get into an argument, and your anger causes you to throw a punch even though you think, at the time of action, that you have most reason to turn the other cheek instead. This is a standard example of a mere behavior: if something inside of you *causes* you to throw a punch against your will, then your throwing a punch does not count as an action but rather counts as a mere behavior. In contrast, I am saying that an agent is akratic when they *act* contrary to what they think, at the time of action, is best for, or from the standpoint of, the person as a whole. So, an analogous case would be: you get into an argument, and you decide to throw a punch because you think that throwing a punch is best for, or from the standpoint of, your current self even though you also think that turning the other cheek is best for, or from the more comprehensive standpoint of, you as a person. That said, even though akratic action and compulsive behavior are, on my view, easy to distinguish in theory, they might not always be easy to distinguish in practice, since it can often be hard to tell, even from the inside, whether the “part of me” that causes a particular action is my current self or a psychological drive that my current self is unable to overcome. And

problematically, many of us might be inclined to opt for the latter kind of self-interpretation more often than we should, especially if we would rather not take responsibility for some of our behaviors.

Third, Holton (1999) and Jackson (1984) say that an agent is weak-willed when they act contrary to what they thought, *before* the time of action, would be best a bit too easily.¹⁰ For example, you might decide to have a drink because you think, *at* the time of action, that having a drink is best even though you thought, *before* the time of action, that staying sober would be best. This is a standard example of *incontinence* (in the philosophical sense of the term): if you think that a fleeting desire to have a drink is reason enough to abandon your general commitment to sobriety, then your having a drink counts as an incontinent action. My account of *akrasia* is very similar to this, but I am adding a twist: I am saying that an agent is *akratic* when they act contrary to what they think, *at* the time of action, is best for, or from the standpoint of, the person as a whole – where their conception of the person as a whole draws from their *past* as well as their *present* beliefs, desires, preferences, etc. So, an analogous case would be: you decide to have a drink because you think, *at* the time of action, that having a drink is best for, or from the standpoint, of your current self even though you also think, at the time of action, that staying sober is best for, or from the standpoint of, yourself as a person – where this latter judgment draws from your past preference to stay sober as well as from your present preference to have a drink. Thus, *akrasia* and *incontinence* are, on my view, related, and many actions might count as both at the same time. However, not all incontinent actions are *akratic* (since you might act contrary to what you thought, *before* the time of action, is best without acting contrary to what you think, *at* the time of action, is best for, or from the standpoint of, yourself as a person), and not all *akratic* actions count as incontinent (since you might act contrary to what you think, *at* the

time of action, is best for, or from the more standpoint of, yourself as a person, without acting contrary to what you thought, *before* the time of action, is best easily enough for your action to count as incontinent).

Finally, Davidson (1969/2001) and Bratman (1979) say that an agent is akratic when they do what they think they have *some* reason to do rather than what they think they have *most* reason to do at the time of action. More specifically: they say that the akratic agent bases their action on an “all out” judgment (of the form “*x* is better than *y*,” for Davidson, or “I shall do *x* instead of *y*,” for Bratman), and they base their “all out” judgment on a *less* comprehensive prima facie evaluative judgment (of the form, “*x* is better than *y* in light of reasons *r*”) rather than on a *more* comprehensive prima facie evaluative judgment (of the form, “*y* is better than *x* in light of reasons *r*’,” where *r*’ is more comprehensive than *r*). For example, you might decide to watch a movie because, even though you think you have *most* reason to grade papers at the time of action, you think you have at least *some* reason to watch a movie at the time of action. My account of akrasia is once again very similar. I agree with Davidson and Bratman that the akratic agent is acting on only *some* of the reasons that they attribute to the person as a whole at the time of action. The difference is that, whereas Davidson and Bratman are saying that the akratic agent’s conception of the person as a whole is a conception of who this person is and why they do what they do *at the time of action* (and is therefore a conception of what I am calling *your current self*), I am saying that the akratic agent’s conception of the person as a whole is a conception of who this person is and why they do what they do *on average* or *in general* (and is therefore a conception of *a temporally extended group that your current self is part of*). Thus, whereas Davidson and Bratman are saying that the akratic agent is acting on only *some* of the reasons that they attribute to themselves at the time of action, I am saying that the akratic agent is

acting on *all* of the reasons that they attribute to themselves at the time of action while acting on only *some* of the reasons that they attribute to a temporally extended person that they are part of at the time of action.

This might not seem like a very big difference. But for our purposes here it makes all the difference in the world, for three reasons. First, whereas Davidson's and Bratman's account of akrasia requires us to sever the link between action and best judgment, my account does not: it says that akratic actions are based on best judgments, and therefore we can accept that akrasia is possible *whether or not* we accept that every action is based on a best judgment. Second, whereas Davidson's and Bratman's account of akrasia fails to explain the confusing, even paradoxical, phenomenology of at least some akratic actions – according to which we experience ourselves as having *most* reason to do what we have only *some* reason to do – my account of akrasia fully explains this phenomenology. Specifically, it says that, when we have this kind of experience, we are thinking about our reasons from two different, and conflicting, epistemic and practical standpoints at the time of action: the standpoint of the person as a whole (which is more comprehensive but less directly action-guiding for us at the time of action) and the standpoint of our current self (which is less comprehensive but more directly action-guiding for us at the time of action). Finally, whereas Davidson's and Bratman's account of akrasia implies that akratic actions are necessarily irrational, my account does not. If you rationally make different best judgments for, or from the standpoint of, your current self and the person as a whole at the time of action, then, all else equal, you are rationally required to act on the former, less comprehensive best judgment rather than on the latter, more comprehensive best judgment.¹¹

5. The Normativity of Akratic Action

I will now close by briefly considering some of the normative implications of my account of akrasia. Specifically, I will argue that, even though akratic actions are not necessarily irrational on my view, they might still be irrational and/or blameworthy in at least four ways.

First, even though we might *sometimes* be rationally permitted to make different best judgments for “part of me” and “me” at the time of action, most of us will usually *not* be rationally permitted to do this. Why? Because most of us have many desires and preferences whose satisfaction requires that we compromise and coordinate with our other selves, and therefore we have strong prima facie reason to think and act from a practical and epistemic standpoint that represents a fair compromise among them. Thus, for example, even if your work self wants to take the promotion *all else equal*, they might also want to compromise and coordinate with your family self more than they want to take the promotion. And if so, then they should think as follows: “Well, the part of me currently in control has *some* reason to take the promotion on the grounds that they prefer to do that all else equal. But they have *most* reason to keep their current job on the grounds that *I* prefer to do that and they prefer to do what I prefer to do all things considered.”

Second, and relatedly, if we accept the equal weight view about peer disagreement, then we should think that, while we might *sometimes* be rationally permitted to make different best judgments from the standpoint of “part of me” and “me” at the time of action, most of us will usually not be rationally permitted to do this either.¹² After all, our selves are likely to be epistemic peers with respect to most issues. Thus, if we accept the equal weight view about peer disagreement, then we should think that, insofar as our selves disagree about matters of fact, they have strong prima facie reason to reject or at least reduce their credence in their original

judgments – in which case they might have epistemic as well as practical reason to proceed on the assumption that what “I as a person” believe is at least as likely to be true, if not more likely to be true, than what “the part of me currently in control” believes all else equal. Thus, for example, if your family self thinks that prioritizing family over work will make you happier overall and your work self thinks that prioritizing work over family will make you happier overall, and if your family self and work self regard each other as epistemic peers with respect to this issue, then they might have epistemic as well as practical reason to think as follows: “Well, if part of me thinks that prioritizing work over family will make me happier overall and part of me thinks that prioritizing family over work will make me happier overall, then I should probably accept that a balance between work and family will, in fact, make me happier overall – even though it might never seem that way to me at the time.”

Of course, as I said above, there will probably be exceptions to these rules. For example, if your work self wants to take the promotion *more* than they want to compromise and coordinate with your family self (and if they think that they can take the promotion while keeping your family self from retaliating, etc.), then it might be perfectly rational for them to make different best judgments *for* “part of me” and “me” at the time of action. Similarly, if your family self has more information, objectivity, etc. than your work self with respect to what will make you happier overall, then it might be perfectly rational for them to make different best judgments *from the standpoint of* “part of me” and “me” at the time of action. And if and when both of these conditions apply, you will not, on my view, be practically or epistemically irrational for acting akratically.

However, even if you are not *irrational* for acting akratically, you might nevertheless be *blameworthy* for doing so. Consider the analogy with the interpersonal case one last time. If you

have so little interest in compromising and coordinating with your family members that it is often rational for you to act contrary to what the family as a whole thinks is best, then, yes, you might count as *rational* for acting this way, but you will probably also count as *selfish* for doing so. And in the same kind of way, if your drinking self has so little interest in compromising and coordinating with your other selves that it is often rational for them to act contrary to what the person as a whole thinks is best, then, yes, they might count as *rational* for acting this way, but they will probably also count as “*selfish*” for doing so – and your other selves will probably, as in the interpersonal case, be justified in criticizing them (“I hate the side of me that comes out at the bar”) and/or coercing and physically restraining them in order to limit their ability to think and act on behalf of the person as a whole (“If I even *think* about going to the bar, I want you to take my keys away from me”).

Of course, there will probably be exceptions to this rule as well. Just as you can presumably be a good family member without *always* subordinating yourself to the family as a whole, your drinking self can presumably also be a good self without *always* subordinating themselves to the person as a whole. And if this is right, then it should be possible for you to rationally *as well as* ethically act akratically every now and then, as long as your doing so is compatible with your ability to compromise and coordinate with your other selves in general.

This brings us to our fourth and final point: even if you are neither *irrational* nor *blameworthy* for acting akratically, you might nevertheless be irrational in your *interpretation* of your akratic action. Specifically, as I said above, if you construct your general self-narrative by thinking about who you are and why you do what you do *on average* or *in general*, and if you then interpret the person as a whole as the agent of your actions in particular situations, then this self-interpretation will be false. Of course, if and when you act *non*-akratically, this kind of self-

interpretation will not be particularly distorting: you will have done what you thought best for, or from the standpoint of, the person as a whole, and therefore it will not be particularly distorting for you to interpret the person as a whole as the agent of your action. But insofar as you act akratically in my sense, this kind of self-interpretation becomes distorting: you will have acted contrary to what you thought best for, or from the standpoint of, the person as a whole, and therefore this kind of self-interpretation will lead you to think of yourself as having acted irrationally or behaved compulsively when you might not have done that at all. In these cases, then, it would be better for you to interpret your current self as the agent of your action, and then to ask whether your current self was rationally and/or ethically permitted to act this way.

To be clear, these brief remarks are not intended to be the final word on the normativity of akratic action; they are rather intended only to point to the kinds of considerations that will be relevant to that discussion. In the meantime, I have argued that many of us can, and do, rationally (and ethically) act akratically, i.e., we can, and do, rationally (and ethically) (a) attribute mental states to “part of me” as well as to “me,” (b) make different best judgments for, or from the standpoint of, “part of me” and “me” at the time of action, and then (c) act on the former, less comprehensive best judgment instead of the latter, more comprehensive best judgment.

Acknowledgements

Thanks to David Barnett, Justin Clarke-Doane, Dale Jamieson, Colin Marshall, Derek Parfit, John Richardson, Nick Riggle, Sharon Street, and J. David Velleman for valuable discussion and feedback on the issues raised in this paper, and thanks to two anonymous referees at *Philosophical Psychology* for very helpful comments on the penultimate draft.

Notes

1. To be clear, my description of you as having a work self and family self is neutral with respect to whether you actually, metaphysically speaking, have multiple selves. But see Dennett (1992), Flanagan (1996), Rovane (1997), Schechtman (1996), Strawson (2011), and Sebo (manuscript) for theories of the self and/or personal identity according to which at least some human beings do, in fact, have multiple selves. Also, see Radden (2011) for an argument that many theories of the self are compatible with the idea of multiplicity.
2. Throughout this paper, I use ‘they’ as third-person singular as well as third-person plural.
3. For detailed discussion of the psychology of multiplicity, see Rita Carter (2008). And for discussion of some of the philosophical implications of multiplicity, see John Doris (2005) and Gilbert Harman (2000).
4. Jeff Sebo (manuscript).
5. I do not mean to imply here that general and situational self-narratives are the only two kinds we construct. In fact, I think that we construct many other kinds as well. But I am choosing to focus on these two kinds here because I think that they are the kinds that, when they conflict, cause us to perform actions that satisfy all the criteria for akrasia.
6. For a discussion about the evolutionary function of the temporal weighting rule, as well as about research on the temporal weighting rule in different species, see Sara Shettleworth (2009, pp. 211-214).
7. For an extended discussion of these social pressures, see Rita Carter (2008, pp. 22-44).

8. Hare offers other explanations of apparent cases of akrasia as well, but this one is most relevant for our purposes here.

9. Watson adds that, in order for you to count as weak-willed rather than merely compulsive, the following counterfactual has to be true: you *could have* done otherwise if, before the time of action, you had developed “certain normal capacities of self-control” (1997, p. 339). But this will not matter for our purposes here.

10. Specifically, Holton says that an agent is weak-willed if they abandon a “contrary inclination defeating intention” too easily (1999, p. 248), where a contrary inclination defeating intention is an intention to perform a particular action and “to overcome contrary desires that one believes one will have when the time comes to act” (*ibid*, p. 250). Similarly, Jackson says that an agent is weak-willed if they perform an action that they had intended not to perform due to the onset of a desire. On this view, the difference between a weak-willed and strong-willed smoker is that the weak-willed smoker “wants not to smoke if he craves to — although when the time comes, he will want to smoke” (1984, pp. 17-18).

11. Alison McIntyre (1990) argues that akrasia can sometimes be rational because our best judgments can be mistaken, and therefore we can sometimes act contrary to what we *think* we have most reason to do while nevertheless doing what we *actually* have most reason to do. I agree with this claim. And if my argument here is correct, then it shows that akrasia can sometimes be rational even if our best judgments are *not* mistaken as well, since we can rationally make two (correct) best judgments at the time of action and then act on the less comprehensive of these two best judgments.

12. For an argument for the equal weight view, see Adam Elga (2007).

References

- Bratman, M. (1979) Practical reasoning and weakness of the will. *NOÛS* 13:2, 153-71.
- Carter, R. (2008) *Multiplicity: the new science of personality, identity, and the self*. New York: Little, Brown and Company.
- Davidson, D. (1969/2001) How is weakness of the will possible? in *Essays on actions and events*. Oxford: Oxford University Press, pp. 21-42.
- Dennett, D. (1992) The self as a center of narrative gravity, in *Self and consciousness: multiple perspectives*, eds., F. Kessel, P. Cole, and D. Johnson. Hillsdale, NJ: Erlbaum Associates, 103-115.
- Doris, J. (2005) *Lack of character: personality and moral behavior*. Cambridge: Cambridge University Press.
- Elga, A. (2007) Reflection and disagreement. *NOÛS*, 41:3, 478-502.
- Flanagan, O. (1996) Multiple identity, character transformation, and self-reclamation, in *Self expressions*. New York: Oxford University Press, pp. 65-87.
- Hare, R.M. (1952) *The language of morals*. Oxford: Clarendon Press.
- Harman, G. (2000) The nonexistence of character traits. *Proceedings of the Aristotelian Society* 100: 223-226.
- Holton, R. (1999) Intention and weakness of will. *The Journal of Philosophy*, 96:5, 241-262.
- Jackson, F. (1984) Weakness of will. *Mind*, 93:369, 1-13.
- McIntyre, A. (1990) Is akratic action always irrational?, in *Identity, character, and morality*, O. Flanagan and A. Rorty (eds.). Cambridge, MA: MIT Press, pp. 379-400.
- Radden, J. (2011) Multiple selves, in *The Oxford handbook of the self*, S. Gallagher (ed.). Oxford: Oxford University Press, pp. 547-70.
- Rovane, C. (1997) *The bounds of agency*. Princeton: Princeton University Press.
- Schechtman, M. (1996), *The constitution of selves*. Ithaca: Cornell University Press.
- Sebo, J. (2013), The self as a center of psychological gravity. Manuscript.

Shettleworth, S. (2009) *Cognition, evolution, and behavior, 2nd edition*. Oxford: Oxford University Press.

Strawson, G. (2011) *Selves*. Oxford: Oxford University Press.

Watson, G. (1977) Skepticism about weakness of will. *The Philosophical Review*, 86:3, 316-39.