The Self as a Center of Psychological Gravity

Jeff Sebo

Abstract

In this paper I develop and defend a conception of the self that strikes a balance between the simplicity of antirealism and the explanatory power of realism. I begin by critiquing two strategies that antirealists have used to try to strike this kind of balance. These strategies culminate in Dennett's analogy with centers of gravity, the potential of which is only partially realized, I argue, in Dennett's conception of the self as a center of narrative gravity. I then argue that we can fully realize the potential of this analogy, as well as vindicate the idea that the self is simple as well as explanatorily useful, if we say that the self is a center of *psychological* gravity, i.e., an abstract, simple set of psychological dispositions around which our actual, fluctuating psychological dispositions are "evenly distributed," which we posit and aim to describe in our self-narratives for purposes of everyday psychological explanation, prediction, and control.

1. Introduction

Many of us think of the self as a simple, unified, "core" set of psychological dispositions that causes many of our actions and experiences. However, we now have very good reason to reject this conception of the self, since it seems very likely that nothing of the sort actually exists. Instead, the evidence tells us, the vast majority of us have an incredibly messy, complex, and dynamic set of psychological dispositions that causes all of our actions and experiences. And as much as we might like to think that we have a coherent and stable psychological "core" that rests at the center of all of these incoherent and unstable psychological fluctuations and causes many of our actions and experiences, the truth seems to be that these incoherent and unstable psychological fluctuations are, for the vast majority of us at least, all that we have.[1]

This development has led many philosophers to ask: how can we revise our traditional conception of the self in a way that makes the existence of the self compatible with the non-existence of a simple, "core" set of psychological dispositions that causes many our actions and experiences? Most philosophers have taken one of two approaches to answering this question. First, many have taken what I will call *the realist approach*: they have argued that we should locate the self in the messy, complex, and dynamic set of psychological dispositions that we actually have. For example, John Perry (2010) argues that the self is simply a person, Galen Strawson (2009) argues that the self is a "subject of experience-as-single-mental-thing," and J. David Velleman (2006) argues that the self is the psychological mechanism responsible for self-narrativity. Second, many philosophers have taken what I will call *the antirealist approach*: they have argued that we should locate the self in the simple, unified "core" set of psychological

---

[1] For detailed discussion of this empirical research, see Carter (2008). And for philosophical discussion of the implications of this research for character traits, see Doris (2005) and Harman (2000).

dispositions that we tell ourselves we have. For example, Daniel Dennett (1992), Owen Flanagan (1996), and Marya Schechtman (1996) all argue that the self is the protagonist of a self-narrative, i.e., the main character of the simple, unified story that we tell ourselves about ourselves.

The costs and benefits of each approach are straightforward: if we take the realist approach, then we get to say that the self causes our actions and experiences but not that the self is simple and unified. And if we take the antirealist approach, then we get to say that the self is simple and unified but not that the self causes our actions and experiences.

My aim in this paper is to develop and defend a conception of the self that strikes a balance between the simplicity of antirealism and the explanatory power of realism. My strategy will be to take what I think is a very promising idea from the antirealist literature and develop it a bit more fully than antirealists have. The idea is that, if we place enough limits on what kind of narrative can constitute a self, then we can ensure that the *protagonist* of this narrative (i.e. the self) will be enough like us that it can *substitute* for us in explanations of our actions and experiences. And in this kind of way, we can ensure that the self, on this antirealist conception of the self, is simple as well as explanatorily useful. However, while I think that this is a very promising strategy for striking a balance between realism and antirealism, I also think that it is incompatible with the idea that the self is, even in part, the protagonist of a *self-narrative* (i.e. the story that we tell ourselves about ourselves), since our self-narratives can be, and often are, mistaken (no matter how many limits we place on them), and therefore the protagonists of our self-narratives can be, and often will be, explanatorily useless. Thus, if we want to say that the self is simple as well as explanatorily useful, then we should locate the self in a different, though related, kind of narrative: specifically, I will argue, we should say that the self is the *intended* protagonist of a self-narrative, i.e., the simple set of psychological dispositions that we *posit* and

aim to describe in our self-narratives, rather than the *actual* protagonist of a self-narrative, i.e., the simple set of psychological dispositions that we *create* and *actually* describe in our self-narratives.

I will proceed as follows. In section 2, I will critique two strategies that antirealists have used in order to make the self, on their conception of the self, explanatorily useful: the strategy of placing limits on the nature of self-narrativity, and the strategy of introducing an "objective" story about who we are and why we do what we do and saying that the self is the protagonist of this "objective" story *as well as* the protagonist of a self-narrative. These strategies culminate in Dennett's analogy with centers of gravity, the potential of which is only partially realized, I will argue, in Dennett's conception of the self as a center of *narrative* gravity. Then, in section 3, I will argue that we can fully realize the potential of this analogy, as well as vindicate the idea that the self is simple as well as explanatorily useful, if we say that the self is the *intended* protagonist of a self-narrative, i.e., the simple set of psychological dispositions that we *posit* and *aim* to describe in our self-narratives. More specifically: I will argue that we should say that the self is a center of *psychological* gravity, i.e., an abstract, simple set of psychological dispositions around which our actual, fluctuating psychological dispositions are "evenly distributed," which we posit and aim to describe in our self-narratives for purposes of everyday psychological explanation, prediction, and control.

2. Why the self is not, even in part, the protagonist of a self-narrative

I begin by critiquing two strategies that antirealists have used in order to make the self, on their conception of the self, explanatorily useful. The first strategy is to place limits on the nature of

self-narrativity. For example, Flanagan and Schechtman both say that a self-narrative is, by definition, a *non-fictional* story that we tell ourselves about ourselves. And the second strategy is to introduce an "objective" story about who we are and why we do what we do and say that the self is the protagonist of this "objective" story *as well as* the protagonist of a self-narrative. For example, Flanagan makes a distinction between identity from the *objective* point of view and identity from the *subjective* point of view, and he says that both kinds of identity count as constituting a self. Schechtman makes a distinction between the story that *actually* makes sense of our actions and experiences and the story that *we think* makes sense of our actions and experiences, and she says that both of these stories contribute to self-constitution. And Dennett says that the self is a center of narrative gravity, where a center of narrative gravity is a theoretical construct that we *posit* as well as a narrative construct that we *create* in our self-narratives. But I will argue that these strategies all fail for the same reason: they say that the self is, at least in part, the protagonist of a self-narrative, and therefore they imply that the self can be, and often will be, explanatorily useless in the relevant sense.

It will help to begin, as antirealists often do, by considering a crude version of my objection so that we can see why we need to place limits on what kind of narrative can constitute a self. According to this crude version of my objection, if the self is the protagonist of a self-narrative, then I can change what my self is like simply by changing what my self-narrative is like. For example, if I want to make my self, say, *trustworthy*, then all I have to do is tell myself a story according to which I always keep my promises, whether or not I think that I actually do, or will. Similarly, if I want to make my self, say, *Napoleon*, then all I have to do is tell myself a story according to which I am a French military and political leader during the French Revolution, whether or not I think that I actually am, or was. And so on. And if this is right, then

4

it follows that my self, on this antirealist conception of the self, is clearly not up to the task of substituting for me in explanations of my actions and experiences. For example, we can all agree that the following explanation would be absurd: "I broke my promise to pick you up at the airport because the protagonist of my *purely fictional* self-narrative is trustworthy." Similarly, we can all agree that the following explanation would be absurd too: "I have no idea how to speak French because the protagonist of my *purely fictional* self-narrative is fluent in French." And so on. Thus, the crude version of my objection concludes, antirealism about the self renders the self explanatorily useless: if the self can be the protagonist of a *purely fictional* self-narrative, then it would clearly not be rational for us to suppose, as a general principle of folk psychology, that our selves cause our actions and experiences – since this folk psychological principle would generate way too many inaccurate explanations and predictions to be remotely useful in everyday life.

This brings us to the first main strategy that antirealists have used in order to vindicate the idea that the self, on their conception of the self, is explanatorily useful: they place limits on the nature of self-narrativity by stipulating that a self-narrative is, by definition, a *non-fictional* story that we tell ourselves about who we are and why we do what we do. For example, Flanagan and Schechtman both say that a self-narrative is, by definition, a kind of *simple autobiography*, i.e., a story that we tell ourselves about ourselves with the aim of making sense of our actions and experiences in simple yet essentially accurate terms, so that we can explain, predict, and control our own behavior easily and effectively in everyday life.[2] This definition of 'self-

---

[2] For example, Flanagan writes
> [T]he author of a true piece of fiction has many more degrees of freedom in
> creating her characters than we have in spinning the tale of our selves. …
> Biographies and autobiographies are constructs. But if they are good biographies
> or autobiographies, they are non-fictional or only semi-fictional. (1996, p. 73)

narrative' implies that, in order for a story that I tell myself about myself to count as a self-narrative in the relevant, self-constituting sense, I have to try to make the *story* fit the *facts* (while also being simple enough to be useful for purposes of everyday psychological explanation, prediction, and control), and I also have to try to make the *facts* fit the *story*. More specifically: I have to (a) attribute to the protagonist of my story actions that I think I performed and experiences that I think I had; (b) attribute to them a simple, unified set of psychological dispositions that I think makes sense of these actions and experiences; and (c) do what I think they would do in light of this simple, unified set of psychological dispositions.[3] Otherwise, if I (a) attribute to the protagonist of my story actions that *I know* I never performed or experiences that *I know* I never had; (b) attribute to them a simple, unified set of psychological dispositions that *I know* make no sense of the actions and experiences that I attribute to them; and/or (c) perform actions that *I know* they would never perform in light of the simple, unified set of psychological dispositions that I attribute to them, then my story does not count as a self-narrative in the relevant, self-constituting sense, and therefore the *protagonist* of my story does not count as my self.

This reply helps a bit. For example, it rules out the possibility that I can make the protagonist of my self-narrative, say, trustworthy or Napoleon simply by telling myself a purely fictional story according to which I am. But while this reply helps a bit, it does not help enough, since, even if I construct my self-narrative with the *aim* of making sense of my actions and experiences so that I can explain, predict, and control my behavior in everyday life, I can still make factual and interpretive mistakes in the construction of my self-narrative that causes me to

---

Similarly, Schechtman writes that self-narrativity is subject to a "reality constraint" that requires, among other things, that self-narrativity aim at factual and interpretive accuracy (1996, pp. 119-130).

[3] Velleman (2006) develops the idea of making the facts fit the story in much more detail.

fall short of *achieving* this aim. For example, I can (a) attribute to the protagonist of my story actions that *I think* I performed but which, in fact, I never performed; (b) attribute to them a simple, unified set of psychological dispositions that *I think* make sense of the actions and experiences that I attribute to them but which, in fact, make no sense of them at all; and/or (c) perform actions that *I think* they would perform in light of the simple, unified set of psychological dispositions that I attribute to them but which, in fact, they would not perform. Schechtman offers the following example: if a man suppresses his memory of all the times he was mean to his siblings, interprets the rest of his behavior towards his siblings more charitably than he should, and so on, then he might sincerely think of himself as a *loving* sibling even though his actions and experiences make sense only on the assumption that he is a *hostile* sibling (1996, p. 117). And insofar as our self-narratives are factually and interpretively mistaken, the protagonists of our self-narratives, like the protagonists of our purely fictional narratives, will clearly not be up to the task of substituting for us in explanations of our actions and experiences. For example, we should hopefully all be able to agree that the following explanation is absurd: "Tim punched his brother in the face because the protagonist of his *sincere but badly mistaken* self-narrative is a loving sibling." Thus, a less crude version of my objection still applies: even if a self-narrative is, by definition, a non-fictional story that we tell ourselves about who we are and why we do what we do, it would still be irrational for us to suppose, as a general principle of folk psychology, that the protagonists of our self-narratives cause our actions and experiences, since, once again, this folk psychological principle would generate too many inaccurate explanations and predictions to be useful in everyday life.

This last claim is easy to misinterpret, so let me briefly distinguish it from two claims that I am *not* making. First, when I say that the protagonists of our self-narratives are not

explanatorily useful in the relevant sense, I am not saying that the protagonists of our self-narratives play no role in shaping who we are and why we do what we do do. On the contrary, they are likely to play a central role in shaping who we are and why we do what we do. For example, if I think of myself as a great artist, then I might spend years working on my art. And if I spend years working on my art, then I might eventually actually *become* a great artist – thereby rendering my initially mistaken self-narrative accurate (at least as concerns the present and future).[4] But this is not the kind of explanation that the antirealist needs in order to vindicate the idea that the protagonist of my self-narrative is explanatorily useful. What the antirealist needs is an explanation that invokes *the protagonist of my self-narrative* (e.g., a great artist) as the cause of my actions and experiences *whether or not* my self-narrative is accurate, not an explanation that invokes *me* (e.g., a person who thinks of himself as a great artist and who therefore does what he thinks a great artist would do) as the cause of my actions and experiences. And the fact remains that, insofar as our self-narratives are factually or interpretively mistaken, the former kind of explanation will often be absurd, and therefore it will not be rational for us to suppose, as a general principle of folk psychology, that the protagonists of our self-narratives cause our actions and experiences.

Second, when I say that the protagonists of our self-narratives are not explanatorily useful in the relevant sense, I am also not saying that it is irrational for us to think of the protagonists of *our own* self-narratives as the cause of our actions and experiences. Why not? Because generally speaking, it is impossible for us to make a distinction, in practice, between, "I believe that *p*" and "*p*," and therefore we have no choice, in practice, but to think of our own self-narratives as

---

[4] Flanagan has this kind of causal explanation in mind when he says that "the activity of self-representation is partly constitutive of actual identity" in part because "the self as represented has motivational bearing and behavioral effects" (1996, p. 69).

factually and interpretively accurate and, consequently, to think of the *protagonists* of our own self-narratives as the cause of our actions and experiences for purposes of everyday psychological explanation, prediction, and control. I think that this well-known epistemic principle explains, at least in part, why so many philosophers have found it plausible to say that the protagonists of our self-narratives figure in causal explanations, or at least explanations that resemble causal explanations: they have found it plausible to say this because they think about the protagonists of their own self-narratives from the first-person perspective, where the gap between our *representations* and *that which is represented* is harder to see. But of course, as soon as we switch to the second- or third-person perspective, this gap becomes much easier to see, and it becomes much clearer to us that what figures in causal explanations is *that which is represented*, not our (possibly factually and interpretively mistaken) *representations*. Moreover, even if we have no choice, in practice, but to think that our self-narratives are factually and interpretively accurate, we can at least acknowledge that they *might be* mistaken (and in some respects probably are); and therefore we should still be able to acknowledge that it would be irrational for us to suppose, as a general principle of folk psychology (a principle whose accuracy and utility should be preserved even if we switch to the second- or third-person perspective), that the protagonists of our self-narratives cause our actions and experiences.

This brings us to the second main strategy that antirealists have employed in order to vindicate the idea that the self is explanatorily useful: they introduce idea of an "objective" story about who we are and why we do do what we do, and they say that the self is the protagonist of this "objective" story *as well as* the protagonist of a self-narrative, so as to prevent our (possibly factually and interpretively mistaken) self-narratives from fully determining what our selves are like in this kind of way. Since the details are different in each case, I will briefly present, and

reply to, three different versions of this strategy: Flanagan's distinction between actual and self-represented identity, Schechtman's distinction between implicit and explicit self-narrativity, and Dennett's conception of the self as a center of narrative gravity.

First, Flanagan makes a distinction between "actual identity" and "self-represented identity," where actual identity is identity from the "objective" point of view and self-represented identity is identity from the "subjective" point of view. More specifically, Flanagan says that

> The actual full identity of some person is the identity the person really has – the
> person he or she really is. … Actual identity is identity from the objective point of
> view. It is the identity which is normally to some significant degree unknown
> from us but which, according to a useful fiction, we come to see with clarity on
> Judgment Day, when all memories are restored and all distortions are removed.
> (1991, p. 35)

In contrast:

> Self-represented identity … is the conscious or semiconscious picture a person
> has of who he or she is. Self-representing involves taking the intentional stance
> toward ourselves and constructing a model that will give us some leverage in
> terms of explanation, prediction, and control over our own lives. … Of course,
> self-represented identity can fall far short of actual full identity, and often far
> afield as well. (*ibid*, p. 37)

Flanagan also says that we can use the term 'self' to refer to either kind of identity (1996, p. ix). Thus, for example, in the hostile/loving sibling case, we might say that my *self-represented* self is a loving sibling whereas my *actual* self is a hostile sibling.

This conception of the self strikes a balance between realism and antirealism by,

essentially, saying that both are true. Specifically, the idea here is that there are at least two kinds of self: the *realist* self, i.e., the complex set of thoughts, feelings, and habits that *actually* cause many of our actions and experiences, and the *antirealist* self, i.e., the simple set of thoughts, feelings, and habits that *we tell ourselves* cause many of our actions and experiences. On this conception of the self, then, we can say that we have a simple self as well as an explanatorily useful self. But of course, our hope was that we would also be able to find a third, middle ground conception of the self according to which the self is simple *as well as* explanatorily useful. And while the distinction between actual identity and self-represented identity is helpful in many respects, it ultimately tells us nothing about whether we can find this third, middle ground conception of the self. What we need, then, is not merely a *conjunction* of realism and antirealism but rather a single, unified conception of the self that strikes a *balance* between them.

This brings us to Schechtman. Schechtman makes a distinction between implicit self-narrativity, which is a story that we *subconsciously* construct and which *actually* makes sense of our actions and experiences, and explicit self-narrativity, which is a story that we *consciously* construct and which *we think* makes sense of our actions and experiences (1996, pp. 115-6). Schechtman then says that both kinds of self-narrative participate in self-constitution, though perhaps our implicit self-narratives contribute to self-constitution a bit more than our explicit self-narratives do.

At first glance, the concept 'implicit self-narrativity' might seem problematic. After all, there is no guarantee that the stories that we *subconsciously* tell ourselves about ourselves will be factually and interpretively accurate either, so how can an implicit self-narrative be a story that we subconsciously tell ourselves about ourselves *as well as* a story that actually makes sense of our actions and experiences? But this problem goes away when we realize that an implicit self-

narrative is, for Schechtman, not a story that we tell ourselves about ourselves at all but rather

what Korsgaard calls 'rudimentary self-consciousness,' i.e., an experience of the world that

represents some of the objects in our environment as "calling out" to be treated in certain ways,

thereby motivating us to treat those objects in those ways (2011, pp. 100-3). For example, if a cat

experiences a mouse as "calling out" to be chased, and if the cat is thereby motivated to chase

the mouse, then we can say that the cat has an implicit self-narrative whose reconstructed content

includes the claim: "I like to chase mice." Similarly, if I experience my siblings as "calling out"

to be treated with hostility, and if I am thereby motivated to treat my siblings with hostility, then

we can say that I have an implicit self-narrative whose reconstructed content includes the claim:

"I am a hostile sibling."[5] Of course, one might resist the idea that this kind of experience

amounts to a self-narrative, since, one might point out, this kind of experience makes only

implicit reference to its subject. Indeed, Schechtman grants that "this is a somewhat unusual way

to think about 'self-narrative'," but she also claims that "very little really turns on this choice of

words" and that, ultimately, we can use whatever language we like so long as we make it clear

"that the psychological forces constituting identity are dynamic and active … rather than static

and passive" (1996, p. 117). In any event, the upshot for our purposes here is that an implicit

self-narrative is, roughly speaking, an experience of the world that causes, and therefore makes

sense of, many our actions and experiences; and therefore the (reconstructed) *protagonist* of an

implicit self-narrative as, roughly speaking, the realist self – or, as Schechtman puts it, "the

psychological organization from which [our] experience and actions are actually flowing" (*ibid*,

---

[5] As Schechtman says, we "confidently ascribe to [the man] a hostility he denies" because, "even
though his explicit account of his actions does not include phrases like 'I hate my brother,' the
assumption that an underlying hostility is playing a role in constituting his experience and
directing his actions is required to make sense of those explicit features of his life." (1996, p.
116)

p. 115).

If this is right, then we can characterize the difference between Flanagan and Schechtman as follows: Whereas Flanagan makes a distinction between two kinds of *identity* (one of which is objective and one of which is subjective), Schechtman makes a distinction between two kinds of *self-narrative* (one of which is objective and one of which is subjective). And whereas Flanagan says that both kinds of identity constitute a self, Schechtman says that both kinds of self-narrative contribute to the constitution of a single, unified self (though perhaps our implicit self-narratives contribute a bit more to self-constitution than our explicit self-narratives do). Thus, whereas Flanagan merely *conjoins* realism and antirealism, Schechtman develops a single, unified conception of the self that strikes a *balance* between them.

But does this kind of hybrid theory strike the right kind of balance between realism and antirealism? That is, does the idea that our implicit self-narratives and explicit self-narratives both contribute to self-constitution justify the idea that the self is simple as well as explanatorily useful? I doubt it. First of all, we would need to hear much more about these two kinds of self-narrative both contribute to self-constitution before we can even speculate about what the self is like, on this view. (And if the answer is just that, as Schechtman puts it, both kinds of self-narrative "shape the way that a person approaches the world" (*ibid*, p. 117), then what originally seemed like a fully antirealist theory of the self and then seemed like a hybrid theory of the self will have turned out to be a fully realist theory of the self after all.) Second, and more importantly, even if we fill out the details of this hybrid theory in a way that results in a true balance between realism and antirealism, why should we think that this balance will preserve the benefits of realism and antirealism while eliminating the costs? It seems much more likely that this theory will strike a balance between the costs and benefits of both theories. On one hand,

13

insofar as facts about my self depend on the content of my *explicit* self-narrative, and insofar as my explicit self-narrative is factually and interpretively mistaken, explanations that invoke my self as the cause of my actions and experiences will still be useless. And on the other hand, insofar as facts about my self depend on the content of my *implicit* self-narrative, and insofar as my implicit self-narrative is complex, my self will still be complex. Thus, if we want a theory of the self that vindicates the idea that the self is simple as well as explanatorily useful, what we need is not merely a conception of the self that strikes a balance between realism and antirealism, but rather a conception of the self that preserves the benefits of each theory while eliminating the costs.

This brings us, finally, to Dennett. Dennett claims that the self is a "center of narrative gravity," where a center of narrative gravity is a *theoretical construct* that we *posit* as well as a *narrative construct* that we *create* in our self-narratives. Dennett begins his discussion by reviewing some of the features of centers of gravity that, he thinks, are relevant to the analogy, and it will be helpful for us to do the same. So what is a center of gravity? As Dennett describes it, a center of gravity is a theoretical construct that we posit in order to explain, predict, and control the behavior of objects. Here is how it works. We want to know whether a particular object – say, a chair – will tip over. How should we proceed? Well, given that the *actual* story about whether the chair will tip over is incredibly complicated, we find it useful to tell a *simpler, semi-fictional* story about it instead. Specifically, we suppose that the chair has a particular point in space around which its weight is evenly distributed, and then we suppose that the location of this point in space determines whether the chair will tip over. Then we explain, predict, and control whether the chair will tip over in everyday life by thinking about the location of this point in space. This, then, is the chair's *center of gravity*: it is the abstract point in space around

which the chair's weight is evenly distributed, which we posit, attribute causal powers to, and theorize about for purposes of everyday scientific explanation, prediction, and control.

Now, two points are worth noting about centers of gravity here. First, they are purely abstract objects. They have no mass, extension, or other physical properties (1992, p. 103). As Dennett puts it, if we wondered whether centers of gravity are really, say, *neutrinos*, then we would be making a category mistake (*ibid*, p. 107). Secondly, however, centers of gravity are *fully objective* abstract objects in the following sense: their existence and nature do not depend, except indirectly, on any of our beliefs about them. For example, you have a center of gravity whether or not you posit one for yourself. Moreover, your center of gravity is the point in space around which your weight is *actually* evenly distributed, not the point in space around which *you think* your weight is evenly distributed (though your beliefs about your center of gravity might cause you to behave in ways that *affect* your center of gravity, and vice versa). Moreover, even though centers of gravity are different from physical objects in some respects (e.g., they do not have any mass or extension, and they can move from one place to another without moving through all the spaces in between), they are similar to physical objects in at least one very important respect (i.e., they figure in causal explanations, or at least explanations that resemble causal explanations) (*ibid*, p. 104).

What, then, is a center of *narrative* gravity? Dennett discusses all of the features of centers of gravity that are relevant to the analogy, but he never spells out exactly how these features apply to centers of narrative gravity. Instead, he simply says that a center of narrative gravity is a theoretical construct that we posit in order to explain, predict, and control the behavior of agents in everyday life:

The physicist does an *interpretation*, if you like, of the chair and its behavior, and

comes up with the theoretical abstraction of a center of gravity, which is then very

useful in characterizing the behavior of the chair in the future... . The

hermeneuticist or phenomenologist – or anthropologist – sees some rather more

complicated things moving about in the world – human beings and animals – and

is faced with a similar problem of interpretation. It turns out to be theoretically

perspicuous to organize the interpretation around a central abstraction: each

person has a *self* (in addition to a center of gravity. (*ibid*, p. 105)

Dennett then says that the self is also a narrative construct that we create by telling ourselves a

story about who we are and why we do what we do so that we can explain, predict, and control

our own behavior in everyday life:

We are all virtuoso novelists, who find ourselves engaged in all sorts of behavior

… and we always put the best "faces" on it we can. We try to make all of our

material cohere into a single good story. [That] story is our autobiography. [And]

the chief fictional character at the center of that autobiography is one's *self*. (*ibid*,

p. 114)

But if both of these claims are true, then the concept of a center of narrative gravity

conflates, or at least combines, two different concepts – the protagonist of the simple story that

we *posit* in our self-narratives and that *actually* makes sense of our actions and experiences and

the protagonist of the simple story that we *create* in our self-narratives and that *we think* makes

sense of our actions and experiences.[6] Thus, Dennett's conception of the self as a center of

---

[6] In other writings, Dennett is ambiguous between objective and subjective interpretations of the concept of a center of narrative gravity. For example, in "The Reality of Selves," Dennett claims that, just as "physicists appreciate the enormous simplification you get when you posit a center

gravity has two related problems. First, it fails to live up to the analogy with centers of gravity, since a center of gravity is *entirely* a theoretical construct that we posit in our folk physical theories and *not at all* a narrative construct that we create in our folk physical theories. (Compare: if facts about our centers of gravity depended, even in part, on the content of our theories about them, then I would be able to, say, move my center of gravity closer to my head simply by telling myself a sincere but badly mistaken story according to which my head weighed 100 pounds.) Second, and as a result, Dennett's conception of the self runs into part of the same problem as Schechtman's: insofar as it makes facts about the self dependent on the content of a self-narrative, it implies that explanations that invoke the self as the cause of our actions and experiences will be useless. (Compare: If I could move my center of gravity closer to my head simply by telling myself a sincere by badly mistaken story according to which my head weighed 100 pounds, then my center of gravity would clearly not be up to the task of figuring in explanations of whether or not I tip over.)

Is it possible to solve these problems by placing further constraints on the nature of self-narrativity? Probably not. After all, the only way to ensure that our self-narratives will be factually and interpretively accurate enough to be useful would be to stipulate that a self-

---

of gravity for an object," "we hetero-phenomenologists appreciate the enormous simplification you get when you posit a center of narrative gravity for a narrative-spinning human body" (1991, p. 418); but this analogy leaves it unclear whether your center of narrative gravity is, in this passage, the simple caricature of you whose beliefs, desires, intentions, etc. makes sense of your *words and deeds* (which would make it like a theoretical construct that we posit) or rather the simple caricature of you whose beliefs, desires, intentions, etc. makes sense of *everything that you tell yourself about yourself with the aim of making sense of your words and deeds* (which would make it like a narrative construct that we create). Either one of these interpretations make sense of the center of gravity analogy as expressed in this passage, since they both posit the self as the protagonist of a simpler, more coherent interpretation of a complex and sometimes contradictory story.

narrative is, by definition, a story that we tell ourselves about ourselves that not only *aims at* but also *achieves* a very high degree of factual and interpretive accuracy. But this stipulation would imply that very few of us, if any of us at all, even *have* self-narratives in the relevant sense – in which case the protagonists of self-narratives would once again not be particularly useful for purposes of everyday psychological explanation, prediction, and control, since so few of us would even have them to begin with.[7] Thus, if we want to vindicate the idea that the self is simple as well as explanatorily useful, then we have to reject the idea that the self is, even in part, the protagonist of a self-narrative.

3. The Self as a Center of Psychological Gravity

I will now argue that, even though both of these strategies for bridging the gap between realism and antirealism fall short, they also point in the direction of a strategy that might succeed. Specifically, if we want to vindicate the idea that the self is simple as well as explanatorily useful, then we should make a distinction between (a) the *actual* story about who we are and why we do what we do, (b) the simple, non-fictional story about who we are and why we do what we do that we *posit* and *aim* to tell in our self-narratives, and (c) the simple, non-fictional story about who we are and why we do what we do that we *create* and *actually* tell in our self-narratives. And we should say that the self is the protagonist of the second of these stories (i.e.

---

[7] Along these lines, Schechtman says that the "reality constraint" on self-narrativity requires that self-narrativity not only aim at but also achieve at least *some* factual and interpretive accuracy. But she also allows that a self-narrative can be mistaken in a variety of ways. (1996, pp. 119-130) And, if my argument here is right, then we can see why she makes this compromise: if she were to require enough accuracy to make the protagonists of our self-narratives explanatorily useful in the relevant sense, then she would have to accept that very few of us have self-narratives, and therefore selves on her view, at all.

the *intended* protagonist of a self-narrative) rather than of the first story, the third story, or some composite of the first and third stories.

We can flesh out the idea of the intended protagonist of a self-narrative in different ways. My aim here will be to develop one particular interpretation of this idea, based on how I think the analogy with centers of gravity should actually play out. I will then argue that this conception of the self as a "center of psychological gravity" vindicates the idea that the self is simple as well as explanatorily useful, and that it does so by incorporating what I take to be the main virtues of the strategies that we considered in section 2.

Here, then, is how I think the analogy with centers of gravity should actually play out. Just as a *center of gravity* is a theoretical construct that we posit in order to explain, predict, and control the behavior of *objects*, a *center of psychological gravity* is a theoretical construct that we posit in order to explain, predict, and control the behavior of *agents*. Here is how it works. We want to know what a particular agent – say, you – will do. How should we proceed? Well, given that the *actual* story about who you are and why you do what you do is incredibly complicated, we find it useful to tell a *simpler, semi-fictional* story about you instead. Specifically, we suppose that there is a simple, unified set of psychological dispositions around which all your actual psychological dispositions are evenly distributed, and then we suppose that this simple, unified set of psychological dispositions causes many of your actions and experiences. Then we try to explain, predict, and control your behavior in everyday life by thinking about what an agent with this simple, unified set of psychological dispositions would think, feel, and do. This, then, is your center of psychological gravity, as I am understanding it: it is the abstract set of psychological dispositions around which all your actual psychological dispositions are evenly distributed,

which we *posit, attribute causal powers to*, and *theorize about* for purposes of everyday psychological explanation, prediction, and control.

Let me explain what I mean by this spatial metaphor. When I say that one set of psychological dispositions "rests at the center" of another set of psychological dispositions, what I mean is that the former set of dispositions represents a coherent and stable *average of* or *compromise among* the dispositions in the latter set. For example, suppose that you tend to have at least some different, and conflicting, beliefs, desires, character traits, and so on in different contexts. For example, you tend to believe that God exists with a .6 credence about half the time (say, while around your theist friends) and you tend to believe that God exists with a .4 credence about half the time (say, while around your atheist friends). You tend to value work a bit more than family about half the time (say, while around your colleagues) and you tend to value family a bit more than work about half the time (say, while around your family). You tend to keep your promises about half the time (say, while around people who tend to keep their promises to you) and you tend to break your promises about half the time (say, while around people who tend to break their promises to you). In this case, we can say that your center of psychological gravity – the abstract psychological "core" around which your actual, fluctuating psychological dispositions tend to "orbit" – believes that God exists with a .5 credence, values work and family in equal measure, is *somewhat* trustworthy, and so on. (Of course, this example is already an oversimplification of most human beings' actual psychological lives, since most human beings' actual beliefs, desires, character traits, etc. change in very small ways very many times a day. But this is fine: the only point of this example is to illustrate the spatial metaphor. Once we have a grasp of the metaphor, it should be clear, at least in theory, how it plays out in more complex cases as well.)

In short, then, the idea is that, since many of us have an incredibly messy and dynamic set of psychological dispositions, we find it useful to posit a simple, unified set of dispositions that "rests at the center" of our actual, fluctuating dispositions, and to try to explain, predict, and control our behavior in everyday life by thinking about what an agent with this simple, unified set of dispositions would choose to do. "Why do I sometimes feel more like a theist and sometimes feel more like an atheist? Because deep down I feel split between the two." "Should I keep my current job, take a promotion that requires me to spend more time at work, or take a demotion that allows me to spend more time at home? Well, deep down I value family and work in equal measure, so I should probably keep my current job." And so on.

Of course, as these examples show, the supposition that our centers of psychological gravity are the cause of our actions and experiences will not always lead us to explanations of our actions and experiences that are *fully* accurate, predictions about our actions and experiences that are *fully* reliable, or decisions about what to do that *maximize* our preference satisfaction at the time of action. For example, in the example above (which, again, is already oversimplified), the *fully* accurate explanation of your sometimes feeling like a theist and sometimes feeling like an atheist is that you sometimes believe that God exists with a .6 credence and you sometimes believe that God exists with a .4 credence, *not* that you always believe that God exists with a .5 credence. (Indeed, you might *never* actually believe that God exists with a .5 credence.) Similarly, the decision that would *maximize* your preference satisfaction at the time of action, we can suppose, is taking the promotion (if you make the decision while at work) or taking the demotion (if you make the decision while at home), *not* keeping your current job. (Indeed, you might never actually prefer keeping your current job.) And so on. Nevertheless, generally speaking, the supposition that our centers of psychological gravity cause many of our actions and

experiences will at the very least lead us to (a) explanations that are very accurate overall and accurate enough in particular cases for most practical purposes, (b) predictions that are very reliable overall and reliable enough in particular cases for most practical purposes, and (c) decisions that come very close to maximizing our preference satisfaction overall and close enough to maximizing our preference satisfaction in particular cases for most practical purposes. Thus, even though explanations that invoke our centers of psychological gravity as the cause of our actions and experiences will sometimes be somewhat off the mark, we are still justified in supposing, as a general principle of folk psychology, that our centers of psychological gravity cause our actions and experiences – because the benefits of this folk psychological theory (simplicity and *very good* explanatory, predictive, and deliberative power) often, if not always, outweigh the costs (less than perfect explanatory, predictive, and deliberative power).

Granted, insofar as a human being has dissociative identity disorder, borderline personality disorder, bipolar disorder, and other such dissociative disorders, explanations that invoke their center of psychological gravity *as a human being* will not be very useful for purposes of everyday psychological explanation, prediction, and control: instead, the only way that we will be able to make sense of most of their actions and experiences is by attributing *multiple personalities* to them, and then attributing a center of psychological gravity to each of their personalities.[8] But this is fine. Compare: insofar as an object has a very stable physical constitution, we will be able to explain, predict, and control its behavior by thinking about only one center of gravity: the center of gravity of the object as a whole. But insofar as an object has multiple, semi-autonomous parts, we will want to think about the center of gravity of each *part* as well (since these parts might also sometimes move independently of each other). Similarly,

---

[8] For psychological discussion of dissociation, see Haddock (2001) and Steinberg and Schnall (2001). And for a philosophical discussion of dissociation, see Hacking (1998).

insofar as a human being has a relatively stable psychology (as most of us do, despite the many

small psychological fluctuations that we experience everyday), we will be able to explain,

predict, and control their behavior by thinking about only one center of psychological gravity:

the center of psychological gravity of the human being as a whole. But insofar as a human being

has multiple personalities, we will want to think about the center of psychological gravity of each

personality as well (since they might sometimes think, feel, and act very differently than each

other). Thus, the fact of dissociative disorders, far from undermining the rationality of supposing

that centers of psychological gravity cause our actions and experiences, actually *supports* the

rationality of this folk psychological practice, since explanations that invoke centers of

psychological gravity as the cause of our actions and experiences are just as useful in the case of

people with personality disorders as they are in the case of people without them (though they will

have to be a bit more fine-grained in the former case).[9]

Now that we have a handle what centers of psychological gravity are, we can see how

centers of psychological gravity, unlike centers of narrative gravity, fully have both of the

features that we emphasized in centers of gravity in section 2. First, they are purely abstract

objects. They have no mass, extension, or other physical properties. If we wondered whether

centers of psychological gravity are actually, say, *brains or souls* (or *parts* of brains or souls)

then we would be making a category mistake. (This much is true of centers of narrative gravity

as well.) Secondly, however, centers of psychological gravity are *fully objective* abstract objects

in the following sense: their existence and nature do not depend, except indirectly, on any of our

---

[9] In fact, the situation is even more complicated than this, since, even in standard cases, we will sometimes want to think about personality-level centers of psychological gravity *as well as* person-level centers of psychological gravity. And the upshot is that, if a self is a center of psychological gravity, then personalities and people can both have selves at the same time. I develop and defend this idea in much more detail in [reference omitted].

beliefs about them. You can have a center of psychological gravity whether or not you posit one

for yourself. Moreover, your center of psychological gravity is the abstract set of psychological

dispositions around which your *actual* psychological dispositions are evenly distributed, not, say,

the abstract set of psychological dispositions around which your *self-attributed* psychological

dispositions are evenly distributed. Thus, even though centers of psychological gravity are

different from physical objects in some respects (e.g., they do not have any mass or extension,

and the principle of bivalence does not hold about them), they are similar to physical objects in at

least one very important respect, i.e., they figure in causal explanations, or at least explanations

that resemble causal explanations.[10]

The upshot is that we now have *three* stories about who we are and why we do what we

do, the last of which has much of what we have been looking for in the story of the *self*. First, we

have what Flanagan calls actual identity, which is the actual story about who we are and why we

do what we do. This story has the benefit of being fully accurate but it also has the cost of being

too complex to be useful for purposes of everyday psychological explanation, prediction, and

control. Second, we have what Flanagan calls self-represented identity, which is the simple non-

fictional story that we create in our self-narratives. This story has the benefit of being simple but

the cost of being too inaccurate to be useful for purposes of everyday psychological explanation,

and control. Finally, we have what we might call *simplified actual* identity, which is the simple

---

[10] To be clear, my description of centers of psychological gravity is meant to be neutral between realism and antirealism about propositional attitudes. If you think that propositional attitudes are real, then the analogy with centers of gravity is straightforward: the self is a simple abstractum that rests at the center of our actual, fluctuating beliefs, desires, intentions, etc. But if you, like Dennett, think that propositional attitudes are abstracta as well, then the analogy with centers of gravity is a bit less straightforward: the idea will be that we attribute to our *personalities* propositional attitudes that make sense of our actions and experiences in particular contexts, and then we attribute to *ourselves* (i.e. our *selves*) propositional attitudes that make sense of our actions and experiences across contexts – where the latter set of attitudes will, as it happens, rest at the center of the attitudes in the former set. For more on this issue, see Dennett (1989).

non-fictional story that we *posit* in our self-narratives. And this third, middle ground story, rather than merely striking a balance between the first two stories, strikes a balance that preserves the benefits of both of these stories while eliminating, or at least minimizing, the costs, i.e., it has the simplicity of self-represented identity and the explanatory power of actual identity (roughly speaking) and, consequently, it has an everyday explanatory *utility* that both of these other stories lack. Thus, the conception of the self as the intended protagonist of a self-narrative, developed here as the conception of the self as a center of psychological gravity, allows us to say that the self is simple as well as explanatorily useful in a way that none of these other antirealist, realist, or hybrid theories has yet been able to do.

It is worth noting how the two main strategies that antirealists have used to try to strike a balance between realism and antirealism play a role in this theory. First, the strategy of placing limits on the nature of self-narrativity sets up the idea that the self is the *intended* protagonist of a self-narrative, since, if self-narrativity constitutively aims at telling a simple story about who we are and why we do what we do that makes sense of our actions and experiences, then a self-narrative is *correct* (i.e. it successfully describes a *self*) insofar as, and only insofar as, it achieves this aim.[11] Second, the subjective/objective distinction sets up the idea that, while facts about selves, like facts about centers of gravity, might be constrained by the constitutive aim of our folk scientific theories (this is the sense in which we *create* selves through our shared folk psychological practices), they are not constrained by the *content* of our folk scientific theories (this is the sense in which we *discover* selves already existing in the world). And, of course, the analogy with centers of gravity sets up the interpretation of the intended protagonist of a self-narrative that I have developed here, according to which the intended protagonist of a self-

---

[11] For similar arguments about belief and action, see Shah and Velleman (2005) and Velleman (2009).

narrative is a center of psychological gravity. Thus, even though this hybrid theory of the self rejects what many philosophers have thought essential to antirealism – the idea that the self is, at least in part, the protagonist of a self-narrative – it also accommodates many of the points that philosophers have made in support of antirealism at least as well as, if not much better than, the theories that preserve this idea do.

Finally, I have been calling my theory of the self a hybrid theory because, as I have argued here, I think that it splits the difference between realism and antirealism in practice. But how should we classify it in theory? This will depend on how restrictive we want our ontology to be. On one hand, if we think that centers of gravity are real because they figure in causal explanations, or at least explanations that resemble causal explanations, then we should think that centers of psychological gravity, and therefore selves on my theory, are real too, and for the same reason. On the other hand, if we think that centers of gravity are merely "useful fictions" because they are abstract objects with no actual causal powers, then we should think that centers of psychological gravity, and therefore selves on my theory, are merely "useful fictions" too, and for the same reason. Ultimately, of course, whether we choose to classify my theory of the self as realist or antirealist will not change the fact that this theory explains and justifies many of our considered judgments about selves. That said, either one of these classifications would have interesting metaphysical implications. For example, if we choose to classify my theory as realist, and if, as I argue elsewhere, this theory implies that people, parts of people, and groups of people all have selves, then it will follow that the number of selves that exist in the world *vastly exceeds* the number of creatures who exist in the world (though this would not, of course, make any particular self any less, or more, valuable in practice). Conversely, if we choose to classify my theory as antirealist, and if, as I argue elsewhere, this theory explains and justifies a lot of what

personality psychologists are up to better than other theories of the self, then it will follow that a lot of what personality psychologists are up to is, technically speaking, folk psychology (though this would not, of course, make personality psychology any less valuable as a field of study).

As this brief discussion suggests, we will have to develop the concept of centers of psychological gravity, as well as the theory of the self as a center of psychological gravity, in much more detail before we can get a full handle on the metaphysical, epistemic, and practical implications of this theory of the self. In the meantime, I have argued that, if we want to say that the self is simple as well as explanatorily useful, then we should say that the self is the intended protagonist of a self-narrative, i.e., the simple set of psychological dispositions that we posit and aim to describe in our self-narratives. And I have developed and defended an interpretation of this idea according to which a self is a center of psychological gravity, i.e., the abstract, simple set of psychological dispositions around which our actual, complex psychological dispositions are evenly distributed. Further issues about, e.g., who has a self, how many selves there are, and so on, will have to wait for another day.[12]

References

Carter, R. (2008), *Multiplicity*. New York: Little, Brown and Company.

Doris, J. (2005), *Lack of Character: Personality and Moral Behavior*. Cambridge: Cambridge University Press.

Dennett, D. (1989), *The Intentional Stance*. Cambridge: The MIT Press.

----- (1991), *Consciousness Explained*. Boston: Little, Brown and Company.

---

[12] Acknowledgements omitted.

----- (1992), "The Self as a Center of Narrative Gravity," in *Self and Consciousness: Multiple Perspectives*, eds., F. S. Kessel, P. M. Cole, and D. L. Johnson. Hillsdale, NJ: Erlbaum Associates, 103-115.

Flanagan, O. (1991), *Varieties of Moral Personality: Ethics and Psychological Realism*. Harvard: Harvard University Press.

----- (1996), *Self Expressions: Mind, Morals, and the Meaning of Life*. New York: Oxford University Press.

Hacking, I. (1998), *Rewriting the Soul: Multiple Personalities and the Science of Memory*. Princeton: Princeton University Press.

Haddock, D. (2001), *The Dissociative Identity Disorder Sourcebook*. New York: McGraw-Hill.

Harman, G. (2000), "The Nonexistence of Character Traits," *Proceedings of the Aristotelian Society* 100: 223-226.

Korsgaard, C. (2011), "Interacting with Animals: A Kantian Account," in T. L. Beauchamp and R. G. Frey, eds., *The Oxford Handbook of Animal Ethics*. Oxford: Oxford University Press, 91-118.

Perry, J. (2010), "Persons and Selves," *Revue de Metaphysique et de Morale*: 455-73.

Schechtman, M. (1996), *The Constitution of Selves*. Ithaca: Cornell University Press.

Shah, N. and J. D. Velleman (2005), "Doxastic Deliberation," *Philosophical Review* 114(4): 497-534.

Steinberg M. and M. Schnall (2001), *The Stranger In the Mirror*. New York: Harper.

Strawson, G. (2009), *Selves*. Oxford: Oxford University Press.

Velleman, J. D. (2006), *Self to Self*. Cambridge: Cambridge University Press.

----- (2009), *How We Get Along*. Cambridge: Cambridge University Press.