

The Metaphysics of the Narrative Self

Michael Rea
University of Notre Dame
mrea@nd.edu

(Final accepted version; forthcoming in *Journal of the American Philosophical Association*)

Narrative has been credited with all manner of different roles in our lives, from contributing to positive outcomes in the wake of trauma, to helping us make sense of and find meaning in our own actions and other events that make up our lives, to unifying our consciousness and explaining important aspects of our agency. It has also, surprisingly, been identified by influential figures in multiple academic disciplines as playing an important role in the very makeup of human identity, selfhood, and personhood. The most provocative claims about the relationship between narrative and the human person are those that suggest that a person, their identity, or their self is somehow identical to or constituted by narrative.¹ Just how provocative these claims are, however, depends on how the notions of *person*, *identity*, *self*, and the relations among them are understood. It also depends on our understanding of *narrative*.

For purposes here, I will assume that an individual counts as a person only if they persist over time and have at some point in their life at least the capacity for agency and subjective experience. Accordingly, although these terms are by no means synonymous, I will generally treat the term “person” as interchangeable with the terms “agent” and “subject”. I will mostly use the

¹ Cf., e.g., Davenport 2012: 2, where he is identifying what he calls the “Signature Thesis” of the narrative identity framework; Jenson 1969: 151; McAdams 2001: 101; and Schechtman 1996:

term “identity” in two different ways—(i) the way that it seems to be used in terms like “identity theft” and “secret identity”, and (ii) the way it is used in terms like “gender identity” and “identity politics”. I will employ the term “autobiographical identity” to signal when I am using “identity” in the first way, and I will employ terms like “social identity” or “gender identity” to signal when I am using “identity” in the second way. Both of these ways of using the term contrast with a third usage which is most common in the literature on identity over time—namely, *numerical identity*, a term that won’t figure much in the present paper.

I will understand *narratives* to be, roughly, interpretive representations (normally linguistic, but perhaps pictorial or even musical), or representations with a contextualizing or interpretive structure. I don’t have a definition of “narrative” in terms of necessary and sufficient conditions to offer, but I can at least add a bit more precision to the characterization as follows: *A narrative, for the purposes of this paper, is a narratively structured representation of a state of affairs or a sequence of events. A narratively structured representation is a representation whose content is unified by an interpretation (tacit or explicit) that orders the various components of the representation in such a way as to highlight their significance in relation to some particular collection of interests (often but not always the interests of the storyteller, the protagonist, or the expected audience), or to identify causal or explanatory relations among those components that are salient in relation to some collection of interests, or both. Narrative structure, then, is to a mere list of events, or to an unordered representation of the components in a complex state of affairs, as form is to matter; and a narrative, or story, may be thought of on analogy with a*

hylomorphic compound of the two.² The notion of tacit interpretation that figures into my characterization of narrative structure can be understood on analogy with the sort of tacit interpretation that is often thought to enter into sense perception. Just as cognitive content—beliefs, desires, fears, and so on—may shape and structure the content of a sense experience (so that, e.g., what look like mere smudges on an x-ray to an untrained layperson may look like fractures or tumors or other signs of disease to an x-ray technician), so too it can shape and structure the content of the self-experience that gives rise to our representations of ourselves and the content of whatever “collective experience” gives rise to a collective representation of a person. Thus, just as it is common to speak of the “cognitive penetration” of perceptual experience, so too we might speak of “narrative penetration” of autobiographical and collective representation.

The term “self” has a wide range of uses in philosophy, psychology, and sociology.³ It is often used in philosophical discussion as a synonym for “person” or perhaps “subject of experience”. On this usage, you *are* a self. But there is another use, common in both philosophy

² This characterization resonates with characterizations offered in the literature by both proponents and critics of “narrative self” views. See, e.g., Schechtman 1996: 97 and G. Strawson 2004: 440 – 441. I got the idea for the analogy with hylomorphic compounding from conversations with Katharina Kraus about the hylomorphic conception of the self that she attributes to Kant (see Kraus 2020, ch. 7).

³ See Gallagher 2000, and Strohming, Knobe, and Newman 2017 and, in a more critical vein, Olson 1998.

and the social sciences, according to which persons *have* selves—past selves, future selves, public and private selves, merely possible selves, and even a “true self”—rather than simply *being* their own, single self. But how are we to understand this latter usage, especially in contexts where we are talking about the *narrative* constitution of the self?⁴ What kind of thing could play whatever social and psychological roles selves are supposed to play and also be constituted by narrative?

Remarkably, proponents of narrative conceptions of identity and the self have, for the most part, shown little interest in developing a detailed metaphysics of identities and selves, largely just taking it for granted that we know what we mean when we use those concepts. Thus, what the literature still badly needs, and what I aim to provide in this paper, is a metaphysics of identities and selves that makes precise the distinction between subject and self and that also makes good metaphysical sense of the idea that identities and selves are constituted by narrative. My goal, then, is not to argue for the claim that these entities are narratively constituted, except perhaps in the limited way that comes with presenting an intuitively attractive metaphysical picture. Rather, the goal is simply to develop a metaphysical theory according to which it is true, and thereby show it to be both intelligible and generally in accord with much of what we want to say about our selves and identities.

⁴ I realize, of course, that talk of “the self” is puzzling in a context in which it is taken for granted that persons have multiple selves. Nonetheless, I will proceed on the assumption that such talk is intelligible, and once my full account of identities and selves is on the table, I will be in a position to explain why it is intelligible.

I begin in section 1 with some very brief remarks about the connections between selves, identities, and the question of who someone is. I then turn, in section 2, to a discussion and critique of the theory of selves set forth by Jenann Ismael in *How Physics Makes us Free* (2016). This, in turn, will serve as a springboard for the development of my own account of identities and selves (sections 3 and 4, respectively). In short, I argue that identities are narratively structured representations, some of which are hosted individually and are autobiographical in form, and others of which are hosted collectively and biographical in form. These identities, in turn, give rise to selves of various sorts—true selves, autobiographical selves, public and private selves, merely possible selves, and so on—which are the characters (or presupposed subjects) that appear in our various identities. Although the theory I develop bears some obvious affinities with the view that selves are *fictional* characters, the two views are in fact distinct, for reasons I explain in section 4.

1. Identity, Narrative, and Who a Person Is

Talk about selves and identities lies in the same conceptual neighborhood as talk about *who* a person is, as contrasted with talk about *what* they are.⁵ When we explain who somebody is, for example, we say something about their identity (or identities). We naturally distinguish between our past self, our present self, our future, possible, superficial, real, and other kinds of selves, and we naturally associate these different selves with representations of who we have

⁵ Cf. Ricoeur 1991.

been, are now, will be in the future, or might be in other circumstances. Moreover, our beliefs about these various selves play a significant role in shaping our emotional states at any given time, our choices, our aspirations, and our sensibilities about the ways in which our life and actions are meaningful; and they do so in precisely the same way and to the same extent as the beliefs we would ordinarily characterize as beliefs about who we are, have been, will be, or could be. Similarly, for our beliefs about the social identities we inhabit—our professional, religious, and gender identities, for example—and about what we might call our “autobiographical identity” (of which more later).

The most thorough sort of answer to the question of who a person is will take a robustly narrative form, and I take this to be at least superficial evidence that narrative is partly constitutive of who a person is.⁶ A mere list of attributes, such as one might find on someone’s CV or in a short biography, or a miscellaneous collection of pictures, such as one might find in a family photo album, or even some combination of these will surely convey *some* sense of who a person is. But the most thorough answers to “Who is...” questions are narratively structured. They contextualize the various roles, attributes, events, so on that they represent, and they provide some clue as to the centrality and significance of these items to the person, to their various communities, or to the audience to which the information is being given.

⁶ This is not to say that the most thorough answer is always the best on any given occasion. Nor is it to suggest that everyone thinks of themselves narratively. (Cf. G. Strawson 2004) The point is just that non-narrative representations will generally be less thorough than narrative ones, even where the two are on a par in terms of the inventory they include.

Compare: A good answer to the question “What happened here?” may or may not tell a story; but a good and suitably full historiographical report of what happened here will tell a story. This suggests a distinction between two senses of “what happened...” questions. In one sense, such questions ask for a mere list of events that occurred at a certain time and location, ones that are salient in relation to contextually determined background interests. In another sense, though, such questions ask for a report of those same contextually salient events that is told through some kind of interpretive lens, one that governs how the author identifies and explains the significance of the events, their contextually salient causal background, the important connections among them, and so on. Corresponding to these two senses of “what happened...” questions are two different ways of understanding the referent of “what happened” in statements of what happened on some particular occasion. For example, understood in the first way, *what happened in the first three decades following Columbus’s arrival in North America* is just a collection of events. It is not something that is constituted by narrative, and, holding fixed the facts about what is salient in the context in which the “what happened?” question is asked, the facts about what happened will be objective—i.e., the facts about what events comprise the relevant collection will be the same for everyone. Understood the second way, however, what happened in those decades will quite literally be a different story for different people, depending on their interpretive lens. In the latter sense of the “what happened?” question, then, *what happened in the first three decades following Columbus’s arrival in North America* is plausibly something that is narratively constituted. So likewise, I submit, for *who a person is* and, accordingly, for their various selves and identities.

As noted in the previous section, I will take it for granted in what follows that selves and identities, whatever exactly those things might be, are both psychologically important and (in a sense yet to be specified) at least partly narratively constituted. I will also take it for granted that the concepts of self and identity are intertwined with the concept of who a person is in the ways I have just described, and that a person's narratively constituted selves and identities include or imply truths about who they are or have been (beyond or instead of mere facts about *what* they are), or about who they can or cannot become downstream of the circumstances in which they are presently situated.

2. Ismael on the Self

As I hope the foregoing has made clear, I am interested here in a conception of self as something *had* by a person, and as the sort of thing that persons may have multiple of, and may have only temporarily (as, e.g., a "past self"). Thus, we can safely rule out views that would identify the self with a substance; for there is no substance, not even a soul, that anyone has more than one of, or only temporarily, that plays anything like the role of self.

Some who endorse narrative conceptions of identity or the self think of identities or selves as events, or collection of events that have intrinsic narrative structure (i.e., meaning relations, facts about significance, explanatory relations, etc. obtain among the events themselves) or some combination of intrinsic narrative structure and further narrative structure

imposed by the mental activity of people reflecting on the events.⁷ Although I do not doubt that there is *some* credible sense in which uninterpreted events or sequences thereof may have intrinsic significance or be structured by intrinsic meaning or explanatory relations, I do not think that such structure deserves to be called *narrative* structure. It is structure that makes the sequence of events *narratable*, of course; but to call it narrative structure is, to my mind, to reduce the provocative-sounding idea that the self is constituted by narrative to the rather mundane and uncontroversial view that the events that constitute a person's life are expressible by way of narrative. More importantly for present purposes, however, is the fact that even if some do want to insist that the sort of intrinsic structure just described is genuinely narrative, it obviously cannot count as narrative structure in the sense relevant to this paper. The reason is simply that, as I am understanding narrative structure, explanation and significance are relativized to the interests of individuals or communities, and the structure inheres in representations and their contents (i.e., propositions or states of affairs) rather than in sequences of concrete events. As something fundamentally interpretive in nature, genuine narrative

⁷ John Davenport (2012) seems to endorse something like the "combination" view just described. (See especially pp. 7 – 9 on the relationship between "practical identities" and what he calls "narratives", and pp. 59, 70 – 71, 89, and 94 on the characterization of narratives and the narrative connections that obtain between the elements in a person's life that are partly constitutive of their narrative.) An anonymous referee asks how exactly my view differs from his. The answer, in short, is that our conceptions of narrative and our ontologies of identities are rather different, as I explain immediately below.

structure is imposed on a representation of a sequence of events, if at all, by minds; and for any sequence of events, there will be as many different ways to narrate the sequence as there are “lenses” of value, interest, and so on through which to view it.

A very different alternative is offered by Jenann Ismael (2016), whose understanding of the self I ultimately reject but will borrow elements from in developing my own. I do not think that Ismael’s view has any claim to being one of the “standard options” in the literature. So far as I am aware, she is its only defender.⁸ But I discuss it here because, despite important drawbacks (which I explain below), the core of her theory is a foundation well worth building on. In the remainder of this section I sketch her view and explain why I cannot accept it; then, in the next section, I highlight what I take to be the core of her theory and proceed to build upon it.

According to Ismael, the (narratively constituted) self is the “grammatical subject of an inner voice”. (71, n. 1) On her view, complex information-handling systems sometimes develop internal “points of view” that allow them to integrate potentially competing streams of information in a way that allows them explicitly to ascribe beliefs, plans, desires, and other intentional states to themselves. Governments and corporations are like this, on her view. Through the pronouncements of its authoritative spokespersons, a government or a corporation itself (not the spokesperson) is able to ascribe intentional states to itself. Certain kinds of brains, or organisms, are like this as well. When a system develops to this point, says Ismael, it “has a

⁸ She claims that her view is broadly Kantian; but one might doubt this. For two alternative interpretations of Kant on the self, see P. F. Strawson 1966 and Kraus 2020.

voice”; and the voice is (so far as I can tell) just the linguistically expressible content of the relevant internal point of view.

But what exactly is this “internal point of view”? As she characterizes them, points of view, or perspectives, are representations whose contents are implicitly relativized to a point or collection of points. The internal point of view associated with the self, a person’s autobiographical point of view, is a *transcendent* viewpoint, one that includes and integrates the contents of the various lower-level points of view that collectively make up one’s life. Insofar as it is a representation whose content is implicitly relativized to a certain point or collection of points, the linguistic expression of the content of one’s autobiographical perspective—the *voice* of that perspective—would naturally be first-personal and narrative in form. But, on Ismael’s view, since there is no object that occupies the perspective, no material or immaterial substance that the contents of one’s autobiographical perspective are in fact relativized to, the “subject” of the inner voice is merely grammatical. As she puts it:

The construction of the point of view literally gives rise to the existence of a subject that occupies it. The subject...isn’t a material object, but something that acts grammatically as the bearer of a point of view when we transform from a perspectival representation into a nonperspectival one.⁹

Suppose for the moment that it is clear what Ismael means by her talk of “something that acts grammatically as the bearer of a point of view”. It is, in that event, relatively straightforward how the self is narratively constituted. The self is not, on this view, made up of narrative; nor

⁹ Ismael 2016: 60.

does the self *itself* have a narrative structure. Rather, one's self comes into existence with and depends for its existence upon the narratively structured content of the representation that is one's autobiographical point of view. In fact, the relation between the self and the content of one's autobiographical point of view seems directly analogous to the relationship between a fictional character and the narrative in which he or she is embedded.

That said, however, I think that it is ultimately not clear what it is to be "something that acts grammatically as the bearer of a point of view". Her claim (quoted above) that the construction of the autobiographical point of view "literally gives rise to the existence of a subject that occupies it" rules out an eliminativist interpretation according to which discourse about the self is really about underlying facts that don't involve any particular subject of experience.¹⁰ She also explicitly rules out the idea that the "grammatical self" (this is my term, not hers) is a material thing; and the idea of an immaterial substance seems not even to be on her radar. Earlier I suggested that the self's dependence upon the content of an autobiographical point of view seems analogous to the dependence of a fictional character upon the narrative in which it is embedded. And some philosophers want to say that the self *is* a kind of fictional character.¹¹ Might we then say that the grammatical self is whatever kind of thing fictional characters are? The trouble with this suggestion is that Ismael's discussion strongly suggests that the self is a mere grammatical subject, whereas this does not seem to be the case for fictional characters.

¹⁰ Cf., e.g., P. F. Strawson 1959: 210 - 212, 1966: 163 – 169.

¹¹ More on this below.

Consider Bilbo Baggins. Either Bilbo exists as a fictional character, or there is no such thing as Bilbo. If the former, then Bilbo is more than a grammatical subject; he or it is some kind of abstract object (presumably) that is genuinely referred to by the term “Bilbo Baggins”. If the latter, then there is no referent of the name “Bilbo Baggins” available to be a grammatical subject. At best, the sentence “Bilbo Baggins is a grammatical subject” will admit of some true paraphrase that does not use the term “Bilbo Baggins” in referential position. So too, *mutatis mutandis*, for the self, if we are to understand the grammatical self on the model of fictional characters. But Ismael does not seem to think of the self as a kind of abstractum that is more than a grammatical subject; nor has she even gestured at paraphrases for sentences like “the self acts grammatically as the bearer of a point of view”. Accordingly, it seems that the self, on her view, cannot be a kind of fictional character. But if it is not that, then I am at a loss to see what it could possibly be.

Toward the end of her discussion of the grammatical self, Ismael declares, “There are no metaphysical mysteries here. The subject appears to itself in consciousness as a locus of introspective awareness that subsumes the sensory modalities and momentary perspectives within a life.” (65) But I have to demur. One mystery obviously concerns what sort of thing can “act grammatically” as the bearer of a point of view. But there is another mystery, too: How can something that acts grammatically as the bearer of a point of view—something that, whatever it is, is dependent for its existence upon a representation—*appear to itself in consciousness*? Talk of appearance is suggestive of something capable of experiencing or attending to the appearance. But Ismael’s characterization of the self as the grammatical subject of an inner voice makes it hard to see how the self could be that sort of thing.

This latter point gets to the main reason why I cannot accept Ismael's conception of the self. I think a non-negotiable desideratum for an adequate theory of human persons and selves is that it must make room, either in the concept of person or in the concept of self, for something that has the power to attend to all of the representations and undergo all of the experiences that are attributed to the person or their self. This desideratum is not satisfied by Ismael's theory.

That said, however, I do think that key elements of Ismael's theory can serve as a foundation upon which to build a theory of selves and persons that does satisfy this desideratum. In this paper, as I have already indicated, my focus is on the theory of selves; and I will not be identifying selves with experiencing subjects. But my point here is that, in contrast to Ismael's theory of selves, mine can be embedded in an overall theory of selves and persons that satisfies the desideratum just described. The first step in developing my theory of selves is to highlight some of the central ideas in Ismael's theory and use them in constructing an account of identity.

3. Narrative and Identity

Despite rejecting Ismael's view of the self, I think that the following core ideas in her view can provide a foundation for a view that is viable:

T1: Perspectives, or points of view, are *representations* centered upon a point or collection of points.¹²

¹² Intuitively, it might seem better to construe a point of view as the point upon which such a representation is centered, rather than the representation itself. But on that construal there is pressure to deny that systems *develop* various points of view (including, in the case of person-

T2: A *transcendent* point of view is one that subsumes—that is, includes and integrates the contents of—a collection of lower-level points of view.

T3: For every person, (i) there are transcendent points of view that subsume various other points of view that (putatively) individually represent parts of their life; and (ii) there is, furthermore, a transcendent point of view that subsumes *all* (and only) these other points of view. Let us call the second type of transcendent point of view the person's *autobiographical* point of view. (This is not, of course, to deny that points of view of the first type are also autobiographical in the ordinary sense of the term.)¹³

like systems, autobiographical points of view) and to say instead that a system *always* has as many points of view as there are points upon which representations might be centered. These latter claims are counterintuitive given our ordinary use of the term “point of view”. Nor is it clear how one might develop a distinction between transcendent and non-transcendent points of view (centered on the same point) without treating points of view as representations rather than points upon which representations are centered. But not much hinges on this way of thinking about things. The theory I will develop could, in principle, be recast with reference to *representations centered on (locational) points of view* rather than with reference to *points of view conceived of as representations centered on a locational point*.

¹³ Admittedly, some persons are psychologically fragmented, as in cases of multiple personality disorder. But I take it that even in the most severe such cases, the fragmentation is not so complete as to eliminate an autobiographical point of view that subsumes the points of view

T4: The content of a person's autobiographical point of view is most naturally expressed in the form of a first-person narrative.

What I now propose is that we drop Ismael's idea that there is something that "acts grammatically" as the bearer of the autobiographical point of view and say instead that, just as the representation that is the content of the point of view is implicitly relativized to a point or collection of points, so too it implicitly posits an individual that is the subject of all the experiences, the agent of all of the actions, and the bearer of all of the points of view that it represents and, furthermore, it implicitly posits this individual as its own bearer, or subject. In simpler terms: just as a fiction written in the first person presupposes a fictional character as the referent of its "I" statements, so too an autobiographical point of view implicitly posits a referent for the "I" statements that would comprise its complete linguistic expression.

The question now is what the relationships are between the narrative content of the autobiographical point of view and the implied referent of its "I" statements on the one hand, and the person, her selves, and her identities on the other hand. In the remainder of this section and the next, I focus on how the autobiographical point of view relates to identities (this section) and selves (the next section).

associated with the various "fragments" of the person. The "and only" qualifier I include just to make clear that there is no threat that the transcendent point of view might (as an anonymous referee put it) "assimilate to itself things that others do or changes that unfold in its environment that it did not actually cause...conflating itself with parts of its setting or its interacting others."

I said earlier that the concepts of both self and identity are intertwined with the notion of who a person is; and I said that a good answer to the question of who someone is will tell a story. On the assumption that selves and identities are narratively constituted, it is natural in light of the conceptual connections just mentioned to think that the stories that do the best job of telling us who a person is are at least partly constitutive of their selves or identities; and it is perhaps also natural to think that the narrative content of the autobiographical point of view will include many, maybe all, of these stories. I think that these ideas are on the right track; however, they need to be adjusted in light of three important observations.

First, the stories we tell ourselves are not the only stories that are told about us. Others have partial stories to tell about us; and, as Hilde Lindemann Nelson (2001) and Kate McLean (2015) argue in different ways and from different disciplinary standpoints, our stories and others' stories about us interact, sometimes stand in tension with one another, sometimes compete with one another for acceptance, and typically help shape one another. Furthermore, as both McLean and Nelson also argue, our stories and others' stories about us are significantly shaped by large-scale cultural narratives about our various social identities and other attributes. Accordingly, there are many different stories, or partial stories, that have some claim to reporting the truth about who we are; and there is no obvious reason to think that our own stories about ourselves ought necessarily to be privileged over other people's stories about us. In other words, who we are *to other individuals*, and who we are *in the various social communities we inhabit*, is not necessarily who we are *to ourselves*; and there is no obvious reason to think that who we are, *simpliciter* is captured by stories we tell ourselves rather than by competing stories about us. So

we should not automatically identify the “stories that do the best job of telling us who a person is” with some portion of the narrative content of the person’s autobiographical point of view.

Second, and relatedly, it is important to bear in mind that, for most contexts, the facts about who we are in that context may include truths about which we are ignorant. We are subject to self-deception, blindspots, delusion, and other failings that inhibit us from believing the sober and unvarnished truth about ourselves. Accordingly, the best, most objectively accurate answer to the question of who a person is (in some particular context) will often include attributes that are unlikely to be included in the (present-tense) narrative content of their autobiographical point of view—attributes like *being a narcissist*, *being in denial about her drug addiction*, *being socially unaware*, or *having the virtue of humility*. If this is right, then, again, we cannot unqualifiedly identify the narrative content of a person’s autobiographical point of view as *the* narrative in which her true self is embedded.

It may be tempting to see these complications as reasons to abandon the narrative approach entirely. Why not instead say that the best answer to the “Who is this?” question for a given person in a particular context is just a proposition that expresses the objective, narrative-independent truth about what is important, unique, deeply explanatory, and so on about the person in that context? Or why not embrace the idea mentioned earlier that the events of a human life have intrinsic narrative structure, and then add that one of the constraints on an adequate answer to the “Who is this?” question is that the answer be faithful to and build upon that intrinsic structure? The answer, in short, is that whatever intrinsic, narratively relevant structure a sequence of life events might have will be too explanatorily thin to constitute a suitably robust answer to the question of who someone is, and this in part because it will not be

tailored to any particular set of interests. As I indicated in the first section, the best answers to the question of who someone is tell a story that exposes explanatory relations and facts about meaning and significance that are relativized to some particular set of interests. They tell us who the person is *to* somebody, or to some community, and they do so by way of narrative. This, to my mind, constitutes good reason to continue trying to develop a narrative conception of self and identity.

In the literature in psychology on narrative identity, identities in general seem to be thought of as a mental constructs of some sort—perhaps a representation or concept of oneself, or an internalized narrative about oneself—that bring “a sense of coherence and integration to one’s life, allowing [one] to perceive a sense of continuity through time”. (McLean 2015: 19) On this way of thinking, an identity is a kind of self-interpretation. But, significantly, the psychological literature here is ultimately focused on who we are *to ourselves*, even if, as McLean argues, this is substantially shaped by facts about who we are to others, and the narratives that others produce about us in response to this. So if we adopt this notion of identity, it makes sense in light of the relativity inherent in the concept of who a person is to think of this as a notion of *autobiographical identity*—one that captures who a person is *to herself* rather than to others, or *simpliciter*.

An autobiographical identity is plausibly identified with some portion of the narrative content of a person’s autobiographical point of view. Specifically, it can be identified with the most comprehensive portion of the narrative content of the autobiographical point of view that plays exactly the sort of unifying, self-interpretive role described by McLean. But what kind of narrative will this be? Not just any narrative about a person’s attributes will constitute an

identity; for an identity is supposed bring coherence and integration to the events of one's life in a way that facilitates a sense of continuity over time. And it is hard to see how a representation could play this kind of integrating role unless it were organized in accord with some notion about what values, desires, goals, and preferences are most *central* or most *important* to the person.¹⁴ Thus, as I shall understand it here, an autobiographical identity will be a proposition that includes and integrates information about a person's most central first- and second-order self-centered and others-centered preferences, values, and goals; their various social identities and their significance to them; the events and experiences that have contributed most to defining their own sense of self (and why they have so contributed); and so on.

Like an autobiographical identity, a *social identity* should also be characterized as representing something central to who a person is. But whereas a person's autobiographical identity represents a central part of who a person is to herself, a social identity represents a central part of who a person is to others. I am a tennis player; but being a tennis player is only a small part of who I am to myself or to others in general. Thus, it is not plausibly a social identity of mine. By contrast, the socially significant attributes that place me in specific professional,

¹⁴ This is not to say that the facts about what is most central to a person are somehow out there in the world independently of anyone's narratively structured representations of them. On the contrary, I am inclined to suppose that all facts about a person's most central values, desires, and so on are relativized to the narratives about them; and then it is an open question whether any particular narrative is privileged over others in terms of the authority of its presuppositions on such matters.

racial, and religious categories *are* social identities of mine. These category-memberships are much more central to who I am both to myself and to others.

A social identity, then, is plausibly identified with the content of what we might call a “collective representation” on the part of others that plays a similarly unifying, integrating role in a (presumably tacit) broader collective representation of who we are. I don’t have an account of collective representation to offer; but I can take at least a few steps toward unpacking it. The notion should not be understood as necessarily implying the existence of anything like a group mind in which the representation is instantiated. Rather, it should be understood in accord with whatever one takes to be the best account of collective belief or collective intention. Perhaps such collective “mental states” are best understood as emergent properties of a kind of group mind; or perhaps instead they are best understood reductively. I do not want to take a stand on the metaphysics of (putative) collective mental states; but I do want to assume, as seems quite plausible, that talk of collective representation makes sense and can be understood along the same lines.

One might object that this way of understanding social identities yields the wrong results in some cases. For example, a serial killer whose identity is unknown doesn’t seem to have *serial killer* as one of their social identities; but, if the serial killer is famous under a label like “the Hillside Strangler”, *serial killer* will be one of the person’s most socially significant attributes, it will be *the* dominant socially significant attribute in the largest collective representation of them, and it will (apparently) play the sort of unifying, integrative role in the collective understanding of who the person is that a social identity is supposed to play. As a result, it might seem that the present account of social identity is on the wrong track.

But in fact I think the worry just mentioned can be addressed by attending to a distinction between *de re* and *de dicto* representational content. Take another example. The attribute of being a *superhero* is—or would be if there were such a people—highly socially significant; and, in the *Superman* fiction, this attribute is clearly part of the *de re* collective representation of Clark Kent. But, because it is not generally known that Clark Kent is Superman, no part of any collective representation of him is aptly expressed with a sentence like “Clark Kent is a superhero”, so it is not part of the *de dicto* representational content of any collective representation of him. Thus, being a superhero is indeed a social identity *of the man* who is called “Clark Kent”; it is a central part of who *he* is to many people, even if almost none of them know that the Clark Kent persona is one of Superman’s *personae*. However, because virtually none of the collective representations of him link being a superhero with the name “Clark Kent” or the superficial attributes he displays when he presents himself under that name, this is why it seems that “Superhero is not a social identity of Clark Kent” is true. So likewise, we might say, in the case of the serial killer.

This reply to the serial killer example suggests a further modification to what has thus far been said about social identities. Consider the Clark Kent persona. What is this, exactly? The name “Clark Kent”, of course, refers to the man himself, which is why “Clark Kent is identical to Superman” expresses a truth even if the *personae* are very different. But the *Clark Kent persona* seems just to be another more complex and comprehensive social identity of the man, Clark Kent. In fact, it seems that we tend to speak of *personae* only when someone gets to the point that a complex cluster of social identities and associated personality and behavioral traits takes on a life of its own, as it were, and becomes a social identity in its own right. Thus, what we might say about the case of Clark Kent is that being a superhero is not a social identity *included in* the Clark

Kent identity, but it is an identity included in the *Superman* identity; and because of the latter, it is a social identity of the man himself.

In light of the foregoing, I propose the following account of social identity. Whereas a person's autobiographical identity is a comprehensive, internalized, self-interpretive (and thereby narratively structured) representation of themselves, a social identity of a person is a partial, external, interpretive (and thereby narratively structured) representation of them that is focused on one or more of their socially significant attributes.¹⁵ People share a social identity in a context just when there is overlap in their socially significant attributes and the collective representations of them in those contexts reflect the same interpretation of those attributes. For the most part, social identities will be relatively simple socially salient attributes like gender, race, professional identity, religious identity, and the like.¹⁶ But in some cases they will be highly complex attributes of the sort we are inclined to describe as *personae*. The context relativity and narrative constitution of social identities I take to be obvious carryovers from what I have said in discussion of autobiographical identities and points of view.

If this is right, then there is a straightforward sense in which our identities (both social and autobiographical) are narratively constituted: both are representations with narrative

¹⁵ This view bears some affinity with Ásta's (2018, ch. 6) account of social identities as locations on a social map.

¹⁶ It might seem that a natural consequence of including gender on the list of social identities together with the account just offered is that gender cannot be (as some would want to hold) a *self-conferred* identity. But I disagree. See Rea 2021 for details.

content, and both correspond to attributes (though only our social identities are simple enough to have common names like ‘race’ or ‘gender’). In the next section, I move from this account of identities to an account of selves.

4. Identities and The Self

Identities, on the view I am developing, are narratively constituted by virtue of being identical to the narratively structured content of a representation of some sort—an autobiographical point of view in the case of one’s autobiographical identity, or a collective representation in the case of one’s social identities. Selves, by contrast, are, intuitively speaking, more like characters *in* a narrative rather than narratives in which characters are imbedded. When I speak of my past self, my future self, my possible selves, or my real self, what I have in mind is something much more like a literary character—a character that captures who I have been, will be, could be, or am deep down, beneath superficial appearances. Not only is this way of conceptualizing selves latent in our ordinary ways of speaking about the self, but it seems to be the dominant view among theorists who embrace what we might call “fictionalist narrativism” about selves—the view that (i) selves are narratively constituted, and (ii) selves are fictions, or fictional characters.¹⁷ If fictionalist narrativism about selves is correct, then selves are narratively constituted in a way different from that in which identities are. Instead of being representations

¹⁷ See Dennett 1989, Gazzaniga 2009, Camp 2011.

structured by narrative, they are—like fictional characters—either parts or creative products of narratives.

On the view I advocate, selves are embedded in identities in the way that fictional characters are embedded in their home narratives. As with fictional characters, selves may be more or less fully drawn; and just as some fictions fail to generate anything robust enough to constitute a fictional character, so too some identities will be too thin to generate a self. Philosophical thought experiments, for example, often take narrative form and sometimes attach names to the hypothetical individuals involved in the experiments. But it is doubtful that any of these individuals are fully-drawn enough to constitute fictional characters. So likewise, with many of our social identities. Collective representations that simply mark someone as a woman or as Pacific Islander, for example, would not be robust and independent enough to generate “characters” distinct from whatever character lies at the center of the most comprehensive collective representations of her, and so they will not generate distinct racial or gendered selves.

Suppose we embrace the view that selves are something like fictional characters that are created by, or embedded in, narratively constituted identities. We might then identify our autobiographical self with the character at the center of our autobiographical identity. We might also, in some cases, identify one or more additional selves of a person with characters embedded in the sorts of complex social identities that I have called *personae*. So, for example, Clark Kent and Superman might be public selves of the same man; and his Kryptonian identity as Kal El might be one of his private selves. Because these selves are identical with various narratively constituted identities, they too will count as narratively constituted.

What of a person's *true self*, the self that corresponds to who the person *really is* (as contrasted with who they *take* themselves to be, or who others take them to be)? *Prima facie*, there is much to be said for identifying a person's true self with their autobiographical self. Ordinarily, each of us has much better access to facts about our past and present actions, about our interests, beliefs, and motivations, about the experiences we have undergone, and so on than any other individual or group of individuals. Thanks to this greater access, each of us is generally better positioned to interpret the events of our lives than others are, too; and, as a result, it will generally be a more authoritative representation of who we are than other representations of us. Accordingly, our own autobiographical identity seems a better candidate for representing who we *really are* than any other representation of who we are, and this, in turn, provides some reason for identifying our true self with the character at the center of our autobiographical narrative.

But if we do identify the character at the center of our autobiographical narrative with our true self, then we must find something to say about the problem of blindspots and delusions mentioned earlier. It may be that narcissism, or self-deception, or excessive humility is a part of who a person really is, even if they don't explicitly believe such things about themselves. And if that is right, then it is unclear how a person's true self could sensibly be tied to the content of their autobiographical narrative. I think the solution to this problem is most easily seen by attending again to fiction, and once we have done so, we will be in a position to draw a helpful distinction between the autobiographical self and the true self.

A novel written in the first person point of view can easily portray its protagonist as a narcissist, as subject to self-deception, as excessively humble, etc. An autobiography can do the

same, even if the author does not recognize these attributes in themselves. (Here the autobiographical remarks of a variety of politicians may immediately spring to mind.) In the first case, it is *true in the fiction*, even though it is not explicit in the fiction, that the protagonist is a narcissist, self-deceived, or whatever; and we can cash this out in terms of whatever theory of truth in fiction ultimately seems best to us. Although autobiography is (typically) not fiction, it seems that truth in autobiography will work in much the same way as truth in fiction—e.g., just as what is true in a fiction need not coincide either with what is in fact true or with what the author believes to be true, so likewise what is true in someone’s autobiography need not coincide with actual truth or with what the author believes to be true. Thus, there is no reason to doubt that one’s true self can be characterized by attributes one does not recognize in oneself or might even explicitly deny of oneself.

To illustrate, consider a mundane case of serious mismatch between who someone is to themselves and who they are to others. Suppose Frank is an alcoholic and suppose this is evident to Frank’s co-workers, friends, and family members but not to Frank himself. To others, and as a matter of objective fact, Frank is in the grip of deep self-deception; he has little self-control where alcohol is concerned; he is mismanaging his responsibilities at work; and he is generally unreliable in a variety of domains as a result. To himself, however, Frank is the opposite of all of this. He sees himself as having deep self-understanding and remarkable control where alcohol is concerned, as keeping up with all of his responsibilities at work, and as generally reliable and relied upon for many important things both in his intimate relationships and in his relationships with colleagues. Furthermore, and importantly in this case, the fact that his friends, family, and co-workers view him in the way just described is in no way included in his autobiographical

identity. How, then, can we seriously think that his autobiographical identity *better* represents his “true self” than the social identity grounded in the collective representation of him that is shared among his friends, co-workers, and family?

Assuming this is a genuinely mundane case—i.e., Frank’s delusion is not so deep as to be relevantly like a brain-in-a-vat scenario—Frank will be aware of how much he has had to drink on various occasions, or of gaps in his memories about drinking; he will be aware of deadlines he has failed to meet, and why he has so failed, even if he has made up excuses for his failures; he will be able to read basic social cues, even if he often finds himself explaining away their significance; he will know what alcoholism and self-control are, even if he incorrectly characterizes himself in relation to those concepts; etc. Accordingly, there will be deep tensions, if not outright incoherence, in his autobiographical identity on matters related to his alcohol consumption and its effects.¹⁸ Even if the identity does not outright represent him as seeming

¹⁸ What if there are no such tensions? What if Frank is never held accountable for his behavior, and he so misunderstands other people, his own thoughts and behavior, and anything else that might challenge his self-narrative that no incoherencies are generated? Then, I think, his situation is more like one (discussed below) wherein someone thinks they are Napoleon than a case wherein someone suffers from mere run-of-the-mill self-deception; and in that case I think it is plausible to say that being an alcoholic might *not* be part of who he really is after all. (In fact, in a case involving such a high degree of delusion, it seems likely that his dependence on alcohol, if any, would be wholly unlike any real-world case of alcohol dependence; and so for that reason alone he may not count as an alcoholic.)

to his friends, family, and co-workers to be an alcoholic, it will represent him as drinking a lot (or suffering memory losses in the wake of drinking) and in a way that in fact manifests lack of self-control, as missing deadlines at work, as subject to a variety of negative social cues that he may find otherwise inexplicable, and as having attributes that in fact satisfy the concept of “alcoholic” even if he explicitly tells himself that he is not an alcoholic. Given all of this, what will likely be true in the narrative content of Frank’s autobiographical identity is that he is a self-deceived, unreliable alcoholic who has an inaccurate understanding of himself. This is just the result we should expect in light of the fact that we typically *blame* people who are self-deceived in ways like this, saying that they *should have known* or really *did know* all along, what kind of person they are; and it is likewise the result we would expect if a person’s true self is captured in their autobiographical identity.

Admittedly, nothing here *guarantees* that a person’s autobiographical identity will capture their own narcissism, delusion, self-deception, or whatever. Someone who is sufficiently deluded might well be to themselves someone radically different from who they are to others and, furthermore, who they are to themselves might be grossly unfaithful to the objective facts. For example, Suppose Fred, living in a psychiatric hospital in the 21st century, believes with conviction and great attention to historical detail that he is Napoleon Bonaparte, lives in the 18th century, and is in the midst of grand conquests. Is there any way that my view of the true self can avoid yielding the judgment that the Napoleonic autobiography captures who Fred really is?

Surprisingly, I think that the answer is *not necessarily*, and that this is a feature and not a bug. I do not think that one’s autobiographical identity can successfully represent them as identical to someone they are not in fact identical to; so it will not follow on my account that it

can be part of Fred's true self to be *identical to Napoleon*.¹⁹ But if his delusion is deep enough, a variety of facts about who Napoleon was—his values, desires for conquest, love for France, and the like—may well be part of who Fred really is. Indeed, if the delusion is deep enough, it may even be that *being a great conqueror* is part of who Fred is, just as one might be a great conqueror in *World of Warcraft* even if one is not in real life.

We can now draw a useful distinction between a person's true self and their autobiographical self. Imagine an autobiography re-written in two different ways: one that successfully omits or negates everything included in the original autobiography that is both inaccessible to the autobiography's first-person narrator and in tension with what they explicitly affirm of themselves, and another that makes explicit all of what is implicit in the narrative but beyond the access of the first-person narrator and then explicitly resolves all resulting tensions in favor of what is true in the narrative rather than in favor of what the protagonist affirms of themselves. Let us characterize the first sort of revision as first-person privileged and the second

¹⁹ This is because it is metaphysically impossible for someone distinct from Napoleon to be identical to Napoleon, and I do not think that we are capable of having genuine mental representations of metaphysically impossible states of affairs. Granted, it sometimes seems as if we are representing an impossible state of affairs. Escher's portrait of an ever-ascending staircase, for example, generates such an illusion. But just as our mental representation of the Escher portrait does not *in fact* represent a genuine ever-ascending staircase, so Fred's apparent representation of himself as identical to Napoleon is not *in fact* a representation of himself as identical to Napoleon. (On this, see also Velleman 1996.)

as omniscience privileging. I take it that the content of each of these revisions is included in the original; the revised versions are obtained simply by removing content and rendering explicit other content. So likewise, it seems reasonable to think that the content of a person's autobiographical identity will include a first-person privileged version and an omniscience privileging version; and, if so, we can sensibly identify the person's autobiographical self—who they take themselves to be—with the protagonist of the first version, and their true self—who they really are—with the protagonist of the second version.

Thus, a person's true self is, in effect, the character at the center of their autobiographical identity *as that character would be understood by a perceptive reader of the narrative who wants to stick to the text in forming their understanding of the character, but who is willing to suspect the character of being unreliable in their own self-understanding*. So the hypothetical reader should not be seen as someone aiming to understand the character simply in accord with the character's own self-understanding, nor should they be seen as aiming to speculatively or interpretively fill in gaps that are left open by the narrative. Rather, the hypothetical reader is one whose goal is to flesh out their understanding of the character by attending to what the text *implies* about them, and letting those implications trump the character's own self-understanding wherever there are conflicts.

We are also now in a position to address a terminological matter that, up to this point, might have seemed puzzling. In a context where we are explicitly recognizing that persons have multiple selves and multiple identities, what sense can it make to speak, in the singular, of *the* self? The practice of speaking in this way (without explanation) is commonplace and has, in my view, rendered the multidisciplinary literature on "the self" rather confusing. Typically, I think

that when people talk about *the* self, using the term “self” in the way I have been throughout this paper, what they have in mind is a person’s *true* self.

My account of identities and selves is now complete. To briefly recap, I have argued that identities are narratively structured representations, some of which comprise our own individual autobiographical identities and others of which are collectively hosted, biographical in form, and comprise our various social identities. Some of our social identities are robust enough to constitute personae, or public selves. But presumably the most important “selves” we have are what I have characterized as our “autobiographical self” and our “true self”, which both arise out of what I have characterized as our autobiographical identity. These selves are all “posited subjects”, analogous to characters in fiction but (unlike characters in fiction) more or less realized in the real world. If I were to go on to explain the relationship between our selves, identities, and the persons that we are, my proposal would be that a person’s autobiographical point of view is akin to a Lewis-style *theoretical definition* of the term that serves as the grammatical subject of its linguistic expression, and that the *person* associated with an autobiographical point of view is the best-candidate-referent for the term implicitly defined by that point of view. But developing this proposal in appropriate detail is a task that must be left for another occasion.²⁰

²⁰ An earlier version of this paper was presented and discussed in a meeting of the metaphysics reading group in the Arché Institute at the University of St. Andrews in June, 2019. I am grateful to the audience members on that occasion (especially Aaron Cotnoir and Matthew McGrath) for helpful comments and discussion. I am also grateful to Laura Callahan, Katharina Kraus, Kate McLean, and Sam Newlands for very helpful conversations about the ideas in this paper, and for

Works Cited

- Ásta. 2018. *Categories We Live By: The Construction of Sex, Gender, Race, and Other Social Categories*. New York: Oxford University Press.
- Camp, Elisabeth. 2011. "Wordsworth's Prelude, Poetic Autobiography, and Narrative Constructions of the Self." *Nonsite.Org* 3.
- Davenport, John J. 2012. *Narrative Identity, Autonomy, and Mortality: From Frankfurt and MacIntyre to Kierkegaard*. New York: Routledge.
- Dennett, Daniel C. 1989. "The Origins of Selves." *Cogito* 3: 163–73.
- Gallagher, Shaun. 2000. "Philosophical Conceptions of the Self: Implications for Cognitive Science." *Trends in Cognitive Science* 4: 14–21.
- Gazzaniga, Michael S. 2009. "The Fictional Self." In *Personal Identity and Fractured Selves: Perspectives from Philosophy, Ethics, and Neuroscience*, edited by Debra Mathews, Hilary Bok, and Peter Rabins, 174–86. Baltimore: Johns Hopkins University Press.
- Ismael, J. T. 2016. *How Physics Makes Us Free*. New York: Oxford University Press.
- Jenson, Robert W. 1969. *The Knowledge of Things Hoped For: The Sense of Theological Discourse*. New York: Oxford University Press.
- Kraus, Katharina T. 2020. *Kant on Self-Knowledge and Self-Formation: The Nature of Inner Experience*. Cambridge: Cambridge University Press.
- McAdams, Dan P. 2001. "The Psychology of Life Stories." *Review of General Psychology* 5: 100–122.
- McLean, Kate C. 2015. *The Co-Authored Self: Family Stories and the Construction of Personal Identity*. New York: Oxford University Press.
- Nelson, Hilde Lindemann. 2001. *Damaged Identities, Narrative Repair*. Ithaca: Cornell University Press.
- Olson, Eric. 1998. "There Is No Problem of the Self." *Journal of Consciousness Studies* 5: 645–57.
- Rea, Michael. 2021. "Gender as a Self-Conferred Identity." Unpublished MS.
- Ricoeur, Paul. 1991. "Narrative Identity." In *On Paul Ricoeur: Narrative and Interpretation*, edited by David Wood, 188–200. London: Routledge.
- Schechtman, Marya. 1996. *The Constitution of Selves*. Ithaca: Cornell University Press.
- Strawson, Galen. 2004. "Against Narrativity." *Ratio* 27: 428–52.
- Strawson, P. F. 1959. *Individuals: An Essay in Descriptive Metaphysics*. First Edition edition. Methuen & Co., Ltd.
- . 1966. *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*. New York: Routledge.
- Strohming, Nina, Joshua Knobe, and George Newman. 2017. "The True Self: A Psychological Concept Distinct from the Self." *Perspectives on Psychological Science* 12: 551–60.

comments on more recent versions of it. Finally, I thank the two anonymous referees for this journal for their thoughtful objections and suggestions.

Velleman, J. David. 1996. "Self to Self." *Philosophical Review* 105: 39–76.

Zeimbekis, John, and Athanassios Raftopoulos. 2015. "The Cognitive Penetrability of Perception: An Overview." In *The Cognitive Penetrability of Perception: New Philosophical Perspectives*, edited by John Zeimbekis and Athanassios Raftopoulos. Oxford: Oxford University Press.