Self to Self

Selected Essays

J. DAVID VELLEMAN
New York University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore,
São Paulo, Delhi, Dubai, Tokyo

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521670241

First published 2006
Reprinted 2007

A catalog record for this publication is available from the British Library

Library of Congress Cataloging in Publication data

Velleman, James David.
Self to Self : selected essays / J. David Velleman.
p. cm.
Includes bibliographical references and index.
ISBN 0-521-85429-6 (hardcover) – ISBN 0-521-67024-1 (pbk.)
1. Self. 2. Self (Philosophy) 3. Kant, Immanuel, 1724–1804 – Ethics. I. Title.
BF697.V45 2005
126–dc22 2005008114

ISBN 978-0-521-85429-0 Hardback
ISBN 978-0-521-67024-1 Paperback

Transferred to digital printing 2009

8

Self to Self

> Images of myself being Napoleon can scarcely merely be images of the physical figure of Napoleon.... They will rather be images of, for instance, the desolation at Austerlitz as viewed by me vaguely aware of my short stature and my cockaded hat, my hand in my tunic.[1]

At the end of "The Imagination and the Self," Bernard Williams uncovers a common confusion about the range of thoughts in which the metaphysics of personal identity is implicated. When I imagine being someone else, I can be described as imagining that I am the other person – which sounds as if I am imagining a relation of identity between that person and me, David Velleman. As Williams points out, however, this particular

[1] Bernard Williams, "The Imagination and the Self," in *Problems of the Self* (Cambridge: Cambridge University Press, 1973), 26–45, 43.

From *The Philosophical Review* 105 (1996), 39–76. Reprinted by permission of the publisher. Throughout my work on this essay, I have benefited from numerous conversations with David Hills. I was also helped by a seminar on metaphysics that I taught with Stephen Yablo, and by Steve's comments on several drafts of the essay. Others who provided comments and suggestions include Paul Boghossian, Linda Wimer Brakel, John Broome, Mark Crimmins, Neil Delaney, Cody Gilmore, Sally Haslanger, Tomis Kapitan, Krista Lawlor, Eric Lormand, Thomas Nagel, Lucy O'Brien, Derek Parfit, Jim Pryor, Henry Richardson, Amélie Rorty, Gideon Rosen, Ian Rumfitt, Sydney Shoemaker, and Paul Torek. This essay was presented at the 1994 Chapel Hill Colloquium, with Michaelis Michael serving as commentator; and to the Philosophy Departments of Princeton and Georgetown Universities. It is dedicated to Claudia Kraus Piper.

way of imagining that I am another person is not really about me or my identity with anyone.[2]

If my approach to imagining that I am Napoleon, for example, is to imagine *being* Napoleon, then I simply imagine a particular situation as experienced *by* Napoleon. I imagine the landscape at Austerlitz as seen through Napoleon's eyes, the sounds of battle as heard through his ears, the nap of a tunic as felt by his hand. Although Napoleon doesn't appear in the resulting mental image, he does appear in the content of my imagining, since I am imagining Austerlitz specifically as experienced by him. But I, David Velleman, am absent both from the image and from the content of the imagining: I'm not imagining anything about the person who I actually am.

Since I'm not imagining anything about my actual self, in this case, I'm certainly not imagining a relation of identity between me and Napoleon. Hence this way of thinking that I am or might be a given person doesn't establish the conceivability – much less the possibility – of any identities between persons.

Unfortunately, metaphysical discussions of personal identity have tended to embrace almost any thoughts about who one is or might be,

[2] Some philosophers have debated whether I can in fact imagine a relation of identity between Napoleon and David Velleman. Bruce Aune argues that I can, provided that I disregard "illusion-shattering facts" about Napoleon and me, such as the fact that I am a twentieth-century philosopher and he a nineteenth-century general ("Speaking of Selves," *The Philosophical Quarterly* 44 [1994]: 279–93, 290 ff.). Zeno Vendler takes the opposite view: "In imagining, for instance, being Ronald Reagan, I cannot be imagining the identity of Z.V. with R.R., for it is patently impossible for these two men to be one and the same, and the patently impossible cannot be imagined" (*The Matter of Minds* [Oxford: Clarendon Press, 1984], 105). (For an answer to Vendler's argument, see John Mackie, "The Transcendental 'I'," in *Philosophical Subjects: Essays Presented to P. F. Strawson*, ed. Zak van Straaten [Oxford: Clarendon Press, 1980], 48–61.)

As Eric Lormand has pointed out to me, however, there are many ways to imagine that I am Napoleon, including not only the method described by Williams but also, for example, imagining that Napoleon has been reincarnated as David Velleman, or that he was cryogenically preserved at birth, thawed out in 1952, and handed by the maternity nurses to an unsuspecting Mrs. Velleman. The latter methods would indeed involve imagining the supposedly problematic relation of identity.

The question, then, is not whether I can imagine a relation of identity between Napoleon and David Velleman but whether I am necessarily doing so when I imagine that I am Napoleon. I interpret Williams as offering a negative answer to this question, by describing a way of imagining that I am Napoleon without imagining anything about David Velleman at all. For a discussion congruent with mine, see Simon Blackburn, "Has Kant Refuted Parfit?" in *Reading Parfit*, ed. Jonathan Dancy (Oxford: Blackwell, 1997), 180–202.

including thoughts similar to the imagining analyzed by Williams. For example, when philosophers want to know whether a person would survive a surgical rearrangement of his brain, they tend to ask whether he would antecedently be in a position to anticipate waking up afterwards. The person's anticipation of waking up after the operation could of course be described as the anticipation that he would survive, in the form of the wakening patient; but it might amount to no more than his picturing the recovery room as seen through the eyes of the wakening patient; and this way of expecting to be that patient is strikingly similar to Williams's method for imagining that one is Napoleon.[3]

If I can imagine that I am Napoleon without imagining a Napoleonic identity for my actual self, then maybe I can anticipate that I will wake up in the future without anticipating a future for my actual self, either. Of course, the anticipation that I will wake up in the future is a first-personal thought; but so is imagining that I am Napoleon; and in that instance, the thought's being first-personal doesn't guarantee that it is about me, the thinker. Imagining that I am Napoleon is first-personal, but it is, so to speak, *first-personal about Napoleon*, in the sense that it is framed from Napoleon's point of view. Perhaps the anticipation that I will wake up in the future can be similarly first-personal about a future subject who may or may not be identical with me. If so, then students of personal identity should probably give up their fascination with first-personal anticipation.

Then again, maybe they should give up their fascination with personal identity instead. The appeal of this topic depends largely on its promise to address our concern about what we can look forward to, or what we can anticipate first-personally. If the mode of anticipation that arouses our concern is first-personal in the sense of being framed from the perspective of a future person, rather than in representing the future existence of the anticipator, then that concern should move us to study the psychology of perspectives rather than the metaphysics of persons.[4]

[3] I believe that Williams himself has gone in for this mode of thinking about personal identity. See, for example, "The Self and the Future," in *Problems of the Self*, 46–63.

[4] At the end of *A Dialogue on Personal Identity and Immortality* (Indianapolis: Hackett Publishing Company, 1978), John Perry has one of the interlocutors conclude, "Perhaps we were wrong, after all, in focusing on identity as the necessary condition of anticipation" (49). This possibility is explored by Raymond Martin in "*Having* the Experience: The Next Best Thing to Being There," *Philosophical Studies* 70 (1993): 305–21. It also figures prominently in Paul Torek's *Something to Look Forward To: Personal Identity, Prudence, and Ethics* (Ph.D. dissertation, University of Michigan, 1995). The present essay is an attempt to find a necessary condition other than identity for the mode of anticipation that arouses our future-directed self-concern.

My aim is to argue for this reinterpretation of our self-regarding concern about the future. What matters most, I shall suggest, is not whether the person I now regard as self will survive into the future; it's whether there will be a future person whom I can now regard as self. And whether I can regard a future person as self, I shall argue, doesn't necessarily depend on whether he will be the same person as me; it depends instead on my access to his point of view.[5]

My first step will be to review the work of other philosophers on first-personal thoughts such as "I am David Velleman" (§1). Drawing on this work, I shall analyze the clause "I am Napoleon" as it is used to characterize what I'm imagining in the case described by Williams (§2). My analysis of this case will lead to some further reflections on the nature of first-personal thought (§3); and the resulting account of the first-person will then be applied to memories of what I've experienced in the past (§4) and anticipations of what I will experience in the future (§5). Our desire for a future to anticipate, I shall argue, is a desire for first-personal access to a future point of view. Why we might have this desire is a question that I'll postpone until the final section of the essay (§6).

1. Who I Am

The connection between identity and perspective has been explored suggestively by Thomas Nagel in his discussions of "the objective self."[6]

[5] In arguing that identity is not what matters about our survival, I am of course following Derek Parfit (*Reasons and Persons* [Oxford: Clarendon Press, 1984]). Let me explain briefly how my views are related to Parfit's.

I agree with Parfit that much of our concern about survival is focused on our psychological continuity with future persons rather than our metaphysical identity with them. But I disagree with Parfit about the kind of psychological continuity that matters to us in this regard. As Parfit conceives it, the relevant continuity comprises not only the psychological connections forged by memory, for example, but also connections forged by the mere persistence of a psychological state or trait (205). I shall argue for a narrower conception of the relevant continuity, as comprising only those psychological connections which function like memory in giving us first-personal access to other points of view. At the end of the chapter, I'll point out that my conception of psychological continuity yields different judgments from Parfit's about various cases in which it's questionable whether the subject survives in the sense that matters.

I think that Parfit himself has reason to prefer my conception of psychological continuity to his own. For as I shall argue, we report our access to other points of view by using the first-person pronoun in ways that would naturally cause this continuity to be mistaken for an identity between persons. My account therefore enables me to explain why that which matters in survival might seem to be identity even when it is not.

[6] "Subjective and Objective," in *Mortal Questions* (Cambridge: Cambridge University Press, 1979), 196–213; "The Limits of Objectivity," in *The Tanner Lectures on Human Values, Vol. I,*

One of Nagel's concerns in these discussions is to locate the fact of who he is:[7]

> [H]ow can a particular person be me? Given a complete description of the world from no particular point of view, including all the people in it, one of whom is Thomas Nagel, it seems on the one hand that something has been left out, something absolutely essential remains to be specified, namely which of them I am. But on the other hand there seems no room in the centerless world for such a further fact: the world as it is from no point of view seems complete in a way that excludes such additions; it is just the world, and everything true of TN is already in it. So... how can it be true of a particular person, a particular individual, TN, who is just one of many persons in an objectively centerless world, that he is me?

Nagel is puzzled here by the fact that he cannot incorporate the thought "I am TN" into an objective description of the world. In an objective description, this thought would have to appear without personal pronouns; but without personal pronouns, the thought would simply disappear. So long as Nagel speaks or thinks of TN in strictly impersonal terms, he cannot frame the thought that TN is *him*.[8]

The impossibility of framing this thought impersonally leads Nagel to worry that a description of the world must remain incomplete so long as it remains impersonal. This worry is metaphysical, in that it envisions things for which "the world" might have "room" even though they cannot be described impersonally. Indeed, Nagel's worry cannot be understood other than metaphysically. Nagel never questions the possibility that an objective description of the world might be complete in the sense of containing all of the objectively statable truths; and its omitting some subjectively stated truths could hardly count against its claim to be a complete objective description. What Nagel envisions is that a description containing all of the objectively statable truths might still be incomplete in the sense of failing to describe all of the world, since the world might include features that cannot be described objectively.[9]

ed. S. McMurrin (Salt Lake City: University of Utah Press, 1980), 77–139; "The Objective Self," in *Knowledge and Mind*, ed. Carl Ginet and Sydney Shoemaker (New York: Oxford University Press, 1983), 211–32; *The View From Nowhere* (New York: Oxford University Press, 1986), Chapter IV.

[7] *The View From Nowhere*, 54–55. Note that this is only one of Nagel's concerns in his discussions of the "objective self."

[8] The classic discussion of this phenomenon is John Perry's paper "The Problem of the Essential Indexical," *Noûs* 13 (1979): 3–21.

[9] The belief in a subjective feature of the world constituting *who I am* is like the belief in a tensed feature constituting *when now is*. The analogy has been drawn explicitly by D. H. Mellor in "I and Now," *Proceedings of the Aristotelian Society* 89 (1988): 79–94. For an

Nagel's reason for thinking that an objective description might be incomplete in this sense is that it could never convey the information conveyed in the subjective statement "I am TN." Nagel's metaphysical worry therefore rests on an observation about the informativeness of an identity statement. And the informativeness of identity statements has been studied extensively by philosophers of language since Frege, including some who have focused especially on identity statements involving the first person.[10]

What the work of these philosophers suggests, however, is that "I am TN" can be informative for Nagel without describing any objectively indescribable feature of the world, and hence that its informativeness shouldn't lead to any metaphysical worries. Let me summarize this work briefly, with the help of David Lewis's suggestion that self-locating thoughts like "I am TN" resemble the cartographic legend "This map is here."[11]

Suppose that you visit the battlefield at Austerlitz and find, at the former site of Napoleon's headquarters, a map that bears the legend "This map is here," followed by an arrow pointing to a rectangle in the map's lower left-hand corner. This legend is certainly informative, but what information does it give you?

The informativeness of the legend depends on the fact that its two indexical terms, "this" and "here," pick out their referents in two different ways. The word "here" is assigned a referent by the arrow that connects it to a rectangle on the map. The word doesn't refer to the rectangle itself, of course; if it did, the legend would make the absurd assertion that the map occupies a small rectangle in its own lower left-hand corner. The word

author who believes in such features of the world, see Geoffrey Madell, "Personal Identity and the Idea of a Human Being," in *Human Beings*, ed. David Cockburn (Cambridge: Cambridge University Press, 1991), 127–42.

[10] I shall be drawing especially on John Perry's "Problem of the Essential Indexical" and his "Self-Notions," *Logos: Philosophic Issues in Christian Perspective* 11 (1990): 17–31. (Both papers have been reprinted in *The Problem of the Essential Indexical and Other Essays* [New York: Oxford University Press, 1993].) See also Stephen E. Boër and William G. Lycan, *Knowing Who* (Cambridge, MA: MIT Press, 1986), Chapter 6; and Lycan, *Consciousness* (Cambridge: MIT Press, 1987), 80. The general account of identity statements on which I rely is similar to that offered by P. F. Strawson in *Subject and Predicate in Logic and Grammar* (London: Methuen, 1974), 51–56.

[11] "Attitudes *De Dicto* and *De Se*," *The Philosophical Review* 88 (1979): 513–14, 528. The moral that I draw from this analogy is similar to one drawn from Kant's Paralogisms of Pure Reason, to the effect that "in identifying 'myself' I am identifying no *more* than a point of view upon the world, and not an entity within it" (Roger Scruton, *Sexual Desire: A Moral Philosophy of the Erotic* [New York: The Free Press, 1986], 114.)

"here" refers instead to the region of the battlefield that's represented by the rectangle, that being where the map is actually located.

The map could refer to this region as "here" without the help of an arrow. For example, it might also bear the words "This map was placed here by the Austerlitz Tourist Board." In this inscription, the word "here" would refer directly to the general vicinity of the inscription itself, and so no arrow would be needed to complete the reference. In "This map was placed here by the Austerlitz Tourist Board," however, the word "here" would roughly mean "where you now see it, before your eyes." And the legend "This map is here" doesn't refer to the relevant region as "here" in the same sense. If the legend "This map is here" was displayed with no arrow, and you had to interpret "here" as meaning "here before your eyes," then the legend would give you no new information. You already know that the map is here before your eyes; what you want to know is where that location lies in the representational scheme of the map. Hence the need for the arrow, which secures reference to the map's actual location via the map's representation of it.

Unlike the word "here," the phrase "this map" does pick out its referent as an object before your eyes. If "this map" referred to the map indirectly, via its representation in the map, then the legend would once again become uninformative. Imagine a second arrow, leading from the phrase "this map" to the same rectangle that's indicated by the arrow leading from "here." This second arrow would reduce the map's legend to the trivial statement that a map located in the region represented by the rectangle is indeed located in the region represented by the rectangle.

The legend on the actual map is informative because it refers to the same location in two different ways – once as the location of "this map [before your eyes]" and once as the location that's "here [according to the map]." The legend tells you where the map that you are seeing can be found on the battlefield as seen by the map.

The reason for referring to the same location twice, as seen by you and by the map, is to help you align the map with your self-centered conception of your surroundings. For until you work out this alignment, you can't use the map to find your way around the battlefield.

In touring the battlefield, you will have to be guided by your senses, which give you a representation of the field from your own point of view. Unfortunately, this self-centered representation of your surroundings is incomplete, in that it includes only what you can perceive or remember perceiving. You want to expand it to include regions that you haven't perceived, so that it represents what is over the hillock on your left or

behind the trees up ahead. These regions are represented in the map, of course, but not from the perspective of the perceptual representation by which you must navigate. You therefore need to transfer information from the map's complete, centerless representation of the battlefield to your incomplete, self-centered representation.

In order to transfer information between these representations, you have to know which parts of them are co-referential – which marks on the map refer to which landmarks within your perceptual field. The legend "This map is here" enables you to coordinate these schemes of reference, by showing how both schemes pick out a single landmark, the map itself.[12]

The informativeness of "This map is here" is thus potentially misleading. "This map is here" adds to your knowledge of the battlefield, but not by giving you knowledge about additional features of the battlefield – features that aren't described in the representations that you already have.

All that the legend reports is the map's location, which is already reported twice in your existing representations of the battlefield, once in the map itself and again in your self-centered conception. Hence the legend doesn't inform you by revealing some aspect of the battlefield that's left out of these representations; rather, it informs you by conveying a rule of translation between these representations, thus enabling you to make better use of the information that they already contain. And the legend conveys this rule of translation by demonstrating it, not by stating it. It *shows* you how to translate between these schemes of representation, by using both of them to specify the map's location.

Many different statements could provide this demonstration. What's conveyed by the legend "This map is here" could equally well be conveyed by a different statement, such as "The hillock on your left is here" or "The trees up ahead are here" or – as maps often say – "You are here." All the legend needs to do is identify some location or other within both representational schemes, thus demonstrating how to translate between them.

In showing you how to translate between schemes of representation, the legend offers practical guidance, which you must follow within the self-centered perspective that you occupy as an agent. That's why the

[12] Gareth Evans took this point further, by suggesting that nothing could count as one's objective conception of the world unless one grasped the possibility of correlating it with one's self-centered conception (*The Varieties of Reference* [Oxford: Oxford University Press, 1982], 212).

legend refers indexically to "this map" and literally points to a region within it, picking out both items as they appear in your visual field. A legend that spoke impersonally about how to transfer information between such-and-such a map and so-and-so's visual field would not be helpful – not, that is, unless you could translate it into your personal terms, such as "this map" and "here." For if you are to follow the rule for translating between the perspectives at hand, that rule must be framed from your own perspective, as it is by the legend "This map is here."

Nagel's thought "I am TN" is informative in the same way: it demonstrates, within his conception of the world as centered on "me," how to correlate that conception with a centerless conception of the world, as containing someone named "TN."[13] "I am TN" is informative, then, because it shows how to transfer information between these two conceptions of the world, not because it describes some feature of the world that they have omitted.[14]

2. Who I Might Be

This account of Nagel's self-locating thought helps us to understand cases of projective imagination as well. My being Napoleon is not a feature of the world that's depicted in the mental image by which I imagine that I am Napoleon; it's rather a rule for translating between that image and an objective description of what it depicts. The image represents that I am Napoleon in the sense that it is framed in a self-centered scheme of reference that's centered on NB.

When I speak of a scheme of reference that's centered on NB, I don't just mean, for example, an image of Austerlitz as it looked from a place where NB stood.[15] Entertaining such an image might amount to no more

[13] Here I am considering, with Nagel, why this statement would constitute an informative addition to a complete objective description of the world. Of course, if Nagel's objective conception of the world is *in*complete, then "I am TN" may be informative in other ways as well.

[14] Nagel explains that "I am TN" is informative because it reports "the fact that this impersonal conception of the world, though it accords no special position to TN, is attached to and developed from the perspective of TN" (*The View From Nowhere*, 64). For a critique of Nagel's explanation, see Christopher Peacocke, *Sense and Content: Experience, Thought, and their Relations* (Oxford: Clarendon Press, 1983), 168–69. A different explanation is offered by Zeno Vendler in Chapter VI of *The Matter of Minds*.

[15] For the sake of simplicity, I am going to confine my attention to the visual image involved in my imagining. Some aspects of visual imagery – for example, its perspectival geometry – are better understood than the corresponding aspects (if any) of tactual, auditory, olfactory, or kinaesthetic imagery.

than visualizing Austerlitz as it looked to NB, which is not the same as imagining that I am NB seeing it. An account of imagined seeing must distinguish it from the less ambitious project of mere visualization.[16] Both imaginative projects involve a mental image drawn from NB's perspective. The difference is that only imagined seeing involves, in addition, the thought of that perspective as occupied – and, indeed, as occupied by NB.[17]

A visual image has a perspective because objects are represented in it by regions whose size and placement depend on the angles subtended by those objects at some common point in space. The representational scheme of the image is governed by lines of sight converging at a single vantage point, whose location the image suggests but doesn't depict.

In ordinary vision, this vantage point is occupied by the eyes of the person experiencing the visual image, and the image is presented as the immediate product of this sensory encounter with the depicted scene.[18] Thus, the image has a *centered* scheme of reference because it represents objects as they are intercepted by lines of sight that converge at a single point; and it has a *self*-centered scheme of reference because the point of convergence is thought of as occupied by the image's subject.

Yet the imagination can frame a visual image without the thought that its vantage point is occupied. The result in that case is visualization rather than imagined seeing. The image represents objects as they would appear to a viewer, if one were present, but it doesn't represent them as so appearing to anyone.

Going beyond mere visualization to imagined seeing entails conjuring up, not just a visual image, but also the thought of such an image as being experienced by someone occupying its vantage point and confronting the objects it depicts. Imaginary seeing thus requires an imagined viewer, who is imagined simultaneously as the mind containing the image, so to speak, and as an unseen object located where its lines of sight converge.

[16] This problem is the one that Williams considers in "Imagination and the Self." The solution I offer here is largely his.

[17] Wollheim distinguishes these modes of imagination as "acentral" and "central" ("Imagination and Identification," in *On Art and the Mind* [Cambridge, MA: Harvard University Press, 1974], 54–83). Williams distinguished them by calling the latter "participatory imagery."

[18] I am being deliberately vague in speaking of how an image is "presented." The "presentation" of the image may consist in a preceding or accompanying thought about the image; or in some distinctive phenomenal qualities of the image itself, combined perhaps with beliefs or cognitive dispositions of the subject with respect to such qualities. I hope to remain neutral among these possibilities.

This viewer is posited by the imagination, but he is not pictured: he is simply thought of, as providing the mental environs of the image and the sensorium at its spatial and causal point of origin.[19]

When I think of the image as having a subject, it becomes a way of thinking about that person reflexively, as "self." And to think of a person reflexively, as "self," is also to think of him as "me." If I think of the image as having a particular subject, such as Napoleon, the image becomes a way of thinking about Napoleon as "me," and so it becomes a way of thinking that I am Napoleon.

Let me elaborate for a moment on this notion of a visual image as a way of thinking about someone else as "me." Elaboration is needed because a visual image rarely contains uses of the first-person pronoun: it isn't a way of thinking about the imagined viewer as "me" in so many words, or in any words at all.

In a case of imagined seeing, however, the image is framed to depict things as seen by someone, who is thus introduced in thought as the subject of the image. The image still doesn't present this viewer as one of the objects visible in it; but it does present the viewer invisibly, insofar as it now depicts things as seen by him; and it thereby presents him reflexively, as the subject, in the way that a spoken first-person pronoun presents its speaker.[20]

Although the reflexivity of a mental image doesn't consist in a use of the first-person pronoun, it would occasion a use of the pronoun in the corresponding verbal report. A report of what I'm imagining would of course describe the objects depicted in the image – the field, the smoke

[19] The relation between the subject's role as the bearer of consciousness and his role as owner of the operative sensorium is discussed by Sydney Shoemaker, "Embodiment and Behavior," in *The Identities of Persons*, ed. Amélie Rorty (Berkeley: University of California Press, 1976), 109–37. It is also the implicit topic of Daniel Dennett's "Where Am I?" in *Brainstorms: Philosophical Essays on Mind and Psychology* (Cambridge: MIT Press, 1981), 310–23. Both papers point out that these roles can come apart.

[20] Throughout the chapter, I assume that "first-personal" thought is not necessarily personal, in that it need not involve the concept of a person. Creatures who lack the concept of a person can nevertheless manifest behavior that is to be explained by their having egocentric representations of their surroundings – representations whose content cannot be expressed without the help of first-person pronouns. We cannot explain the stalking behavior of a cat, for example, except in terms of perceptions expressible as "There's a mouse in front of *me*," "*I'm* close enough to pounce on it," and so on. Yet the attribution of such first-personal thoughts to the cat does not imply that it thinks of itself, or of anything else, as a person. Here I am in pointed disagreement with John Campbell, who thinks that even proprioceptions such as "I am about to fall over" are essentially about a person ("The Reductionist View of the Self," in *Reduction, Explanation, and Realism*, ed. David Charles and Kathleen Lennon [Oxford: Clarendon Press, 1992], 380–419, 392 ff.).

of battle, and so on. Yet it would also have to make clear that these objects were being imagined, not merely as they would appear if someone saw them, but as being seen. How could a verbal report make clear that it was conveying the contents of an imagined seeing? The obvious way would be to include a prefatory "I see," in which "I" would refer to the person who does the seeing; and the person who does the seeing, in this context, is the imagined viewer. The verbal expression of an imagined seeing thus confirms that its scheme of representation casts the imagined viewer in the role of first person, as the referent of "me."

But who would be speaking here? Whose image is being put into words?

I have thus far neglected to distinguish between the image that's in the mind of the imaginer and the one that's in the mind of the imagined viewer. When I imagine that I am Napoleon viewing Austerlitz, I don't imagine, of the faint and incomplete image in my own mind, that this very image belongs to a visual experience in the mind of NB.[21] Rather, my image is a medium for imagining NB's visual experience.

My image is a medium for imagining NB's experience because it purports to be a secondary version of NB's visual image – a duplicate of his visual impression, or a prototype for it. And the image regarded as having NB for its subject would seem to be the primary or original image in NB's mind, not the secondary version of it in mine. The question therefore arises whether my image still qualifies as a way of thinking about NB as "me."

By and large, secondary versions of an image share its referential scheme. A reproduction of a picture of Austerlitz is itself an image of Austerlitz; an artist's design for a mural of Austerlitz is an image of Austerlitz, too. Both are copies – one modeled after the primary image, the other serving as a model for it – and both share the referential scheme of the picture to which they stand as copies.[22] Similarly, the image in my mind, regarded as a copy of NB's visual impression, is an image of whatever NB is supposed to be seeing.

But what about reflexive or first-personal reference? In the referential scheme of NB's visual impression, NB occupies first-person position, since he is the subject. Yet the copy occurs in my mind, where I am the subject. So shouldn't I, DV, be the person who is reflexively presented in this image?

[21] Here I disagree with John Mackie's suggestion that the imagined subject is imagined to be "the subject of my present experiences" ("The Transcendental 'I'," 56).

[22] In speaking of mental images as "copies," I do not mean to imply anything about their degree of resolution, detail, or faithfulness to the original. I am also attempting to remain neutral on the direction-of-fit between these copies and their originals.

There isn't a simple answer to this question. A mental copy of a visual impression can have two subjects. The person entertaining a secondary image is certainly the subject of that image. But insofar as the image is regarded as a copy of a primary impression, it resembles that impression not merely in depicting the objects seen but also in depicting those objects *as seen by the primary viewer.* Allusion to the primary viewer is essential to the representational scheme of the secondary image, and he is alluded to specifically as the subject, since objects are represented specifically as seen by him.

Considerations such as these have led some philosophers to speak of secondary images as having an "internal" subject in addition to any "external" subject they might have.[23] I find the terms "internal" and "external" uninformative, however, and so I will speak instead of the notional and actual subjects. The notional subject of a secondary image is the person thought of as occupying the image's vantage point and undergoing the visual impression of which the image is a copy.

In the representational scheme of such an image, the notional subject tends to crowd out the actual subject as the target of reflexive reference. The notional subject has to get into the act somehow, or the image won't amount to a representation of things as seen by him. And he can't get into the act, in his capacity as the viewer, just by getting into the image; for as the viewer, he occupies a role over and above that of anything viewed. He therefore gets into the act by being thought of as the subject, as the person reflexively presented by the image, and hence as the target of self-reference within the visual scheme of representation.

Consider again how the referential scheme of my mental image would be expressed in words. To ask whom the image presents in the position of subject or self is to ask how the image's self-centered scheme of reference is oriented in the objective world.[24] And as we have seen, an image's orientation can be demonstrated within its scheme of reference by an identity statement of the form "I am so-and-so."

[23] The term "internal subject" was coined, I believe, by Richard Wollheim. Wollheim's clearest discussion of the issue is in Lecture III of *Painting as an Art* (Princeton: Princeton University Press, 1987). In the case of paintings, of course, there is no external subject, since the secondary image is on canvas rather than in a person's mind. See also Wollheim's "Imagination and Identification." For a recent discussion of the issue in application to perceptual experience, see Bill Brewer, "Self-Location and Agency," *Mind* 101 (1992): 17–34.

[24] The objective world involved here is the *imaginary* world, objectively described. After all, I can imagine that I am Napoleon at the battle of Narnia rather than Austerlitz.

If such a statement were framed within my image's scheme of reference, it would be framed from the point of view embodied in the image, which is that of the imagined viewer. And a statement framed from the viewer's point of view would be a statement made by the viewer – who has to be Napoleon if my image is to represent things as seen by him. The identity statement that would demonstrate the referential orientation of my image is therefore the statement that would be made by NB: "I am Napoleon."

In his capacity as the viewer, of course, NB is merely imaginary, and his statement would be imaginary, too. But it would be easy enough to imagine. In fact, I may already be putting imaginary words into the mouth of NB, if my imagining includes what Williams calls a "narration":[25]

> Consider now the *narration*. . . . It is going to be of the general form: 'I have conquered; the ideals of the Revolution in my hands are sweeping away the old world. Poor Maria Walewska, I wonder where she is now' and so on and so on, according to whatever knowledge or illusions I possess about Napoleon.

When I imagine saying "I have conquered," I conjure up an image of this utterance from the speaker's point of view, and I superimpose this point of view on that embodied in the imagined visual impression, in such a way that both are centered on NB as the notional subject of speech and vision together. If I replaced "I have conquered" with "I am Napoleon" (or perhaps "I, Napoleon, have conquered"), I would thereby give myself a demonstration, within the referential scheme of my imagining, of how that scheme is coordinated with an objective description of the world.

To imagine saying "I am Napoleon" would therefore be a way for me to spell out for myself that I'm imagining everything as seen (and said) by NB.[26] I could even use this statement to spell out for others what I'm imagining, provided that I enclosed it in quotation marks to indicate that it was couched in the terms of the imagining. For I could say this: I am imagining, "I am Napoleon."

My report of imagining *that* I am Napoleon simply transposes this quoted identity statement into indirect discourse. In doing so, it replaces

[25] "Imagination and the Self," 43.

[26] The imagined statement itself is not what gives my imagining the content that I am Napoleon. For I can imagine saying "I am Napoleon" without imagining that I am Napoleon – for example, in the course of imagining that I am someone with Napoleonic delusions. To imagine that I am Napoleon is to imagine that which this imagined statement would express – namely, Napoleon's occupying the center of a self-centered scheme of reference.

the pronoun "I" with something like what Castañeda called a quasi-indicator.[27]

A quasi-indicator is an indexical used in *oratio obliqua* to mark the position that would be occupied in *oratio recta* by a reflexive term such as "me." John Perry has analyzed the workings of quasi-indicators as follows:[28]

> I think that when we use quasi-indicators we combine a remark about what [someone] believes with a remark, or a hint, about *how* he believes it. In the case of "he," the second bit of information is roughly that he believes what he believes *in virtue of* accepting a sentence with "I" in it. That is, "Smith believes that *he* is α" tells us that Smith believes Smith to be α in virtue of accepting "I am α." More precisely, it tells us that he [believes] it in virtue of being in a certain belief state, which in English-speaking adults typically results in the utterance, in appropriate circumstances, of "I am α."

Suppose that Smith overhears a conversation in which some unnamed person is confidently said to be α. Smith may come to believe, of that unnamed person, that he is α. Now suppose that the person under discussion is in fact Smith. Smith has then come to believe *Smith* to be α. But Smith may or may not be aware of being the person in question, and so in believing Smith to be α, he may believe it in one of two ways, which Perry analyzes as follows. He may believe it either by accepting a sentence of the form "He is α" or by accepting the sentence "I am α," depending on which sentence would typically be uttered by an English speaker in his state of mind.

When we say "Smith believes that he is α," we normally mean that Smith holds his belief in the latter, first-personal way: our report would be misleading if Smith were unaware of being the person in question. According to Perry, then, we mean not only that Smith believes Smith to be α but also that he believes it in virtue of accepting the sentence "I am α" – that is, in virtue of occupying a state that typically results in an utterance of this first-personal sentence. We thus use "he" as a quasi-indicator, marking the presence of a first-person pronoun in the sentence whose utterance would typically express Smith's belief.

As it stands, Perry's analysis applies only to beliefs: it cannot cover cases of imagining, because imaginings don't typically give rise to utterances. But the materials for extending the analysis are already at hand.

[27] See "Indicators and Quasi-Indicators," *American Philosophical Quarterly* 4 (1967): 203–10. A discussion of the literature on this subject can be found in John Perry's "Castañeda on He and I," in *The Problem of the Essential Indexical.*

[28] "Belief and Acceptance," in *The Problem of the Essential Indexical,* 53–67, 60. Note that Perry's account is different from Castañeda's.

For when Smith imagines that he is Napoleon, we have found, he may do so by conjuring up secondary images with NB as their notional subject, thereby entering a state of imagination whose referential orientation would be spelled out by a further image, of the utterance "I am Napoleon." Just as there is an actual utterance by which the believer would typically express what he believes, so there is an utterance-image by which the imaginer would typically express what he imagines. So Perry's analysis can be extended from beliefs to imaginings if the utterances expressive of beliefs are replaced in the analysis by the utterance-images expressive of imaginings.

This extension of Perry's analysis crucially affects the role of the quasi-indicator. In "Smith believes that he is Napoleon," the quasi-indicator "he" marks the place of the first-person pronoun in "I am Napoleon" as it might actually be said by Smith. The quasi-indicator thus stands in for a pronoun referring to Smith. But in "Smith imagines that he is Napoleon," the quasi-indicator marks the place of the first-person pronoun in "I am Napoleon" as it might be *imagined* by Smith but as *said* in this imagining by Napoleon. And in "I am Napoleon" as said by NB, "I" would refer to NB.[29]

Thus, the 'he' in "Smith imagines that he is Napoleon" echoes an imagined use of "I" that would refer to Napoleon and not to Smith. So it does not pick out Smith as the object of Smith's imaginings; it merely introduces the self-concept, or "I," under which Smith imagines Napoleon, as he would express by going on to imagine saying, "I am Napoleon." The same goes for the second occurrence of "I" in "I'm imagining that I am Napoleon." This 'I' isn't a reference to me, David Velleman. It simply marks the place of the first-person pronoun in the utterance-image "I am

[29] Here is a complication. In Perry's example ("Smith believes that he is α"), the quasi-indicator borrows its reference by anaphora to indicate *what* is believed; whereas it invokes the associated utterance only for the purpose of specifying *how* this content is believed. Fortunately, the grammatical antecedent of "he" (namely, "Smith") has the same referent as the pronoun to which it corresponds in the associated utterance ("I"), so that the *what* and the *how* of the attributed belief coincide. Yet if both of these mechanisms were at work in "Smith imagines that he is Napoleon," then what Smith was said to imagine would be something that he couldn't imagine in the way that he was said to imagine it; since the grammatical antecedent of "he," in this attribution, doesn't have the same referent as the pronoun "I" in the utterance naturally associated with the attributed imagining. Thus, the normal mechanisms of quasi-indication no longer work together. What I am suggesting is that, in case of such a conflict, the mechanism peculiar to quasi-indication takes precedence, so that no anaphora occurs, and both the *what* and the *how* of Smith's imagining are determined by the associated utterance. (Thanks to Tomis Kapitan for raising this problem.)

Napoleon," which would demonstrate the orientation of my imagining from within.

Here at last we see why Williams's method for imagining that I am Napoleon does not involve imagining anything about my actual self, DV. It simply involves entertaining imaginary thoughts in the Napoleonic first-person, so to speak, an egocentric scheme of reference whose center – and hence whose *ego* – is NB.

3. What "I" Is

But how can I think about Napoleon in the first-person? The first-person is a reflexive mode of thought, and I am in no position to think about NB reflexively, since reflexive thoughts are about their own thinker, and I, the thinker, am not NB.

I am happy to grant that my thoughts in this case are not reflexive in the objective sense of referring to the person who is in fact thinking them. But as Perry's analysis illustrates, philosophers have had to recognize a distinction between a thought's being objectively reflexive in this sense and its being subjectively reflexive, by presenting the thinker in the distinctively first-personal way, under the guise of self.[30] Although my thoughts about NB aren't about their own thinker, they do present NB in first-personal guise.

I now seem to be suggesting that some modes of thought may be subjectively but not objectively reflexive, presenting first-personally someone who is not the person thinking them. This suggestion would be problematic, to say the least.

Even those philosophers who recognize the distinction between subjective and objective reflexivity assume that a subjectively reflexive mode of thought – though individuated, perhaps, by its subjective character – must nevertheless be guaranteed to refer to the thinker in fact.[31] Otherwise, I could think about someone first-personally and yet be uncertain of his relation to the thinker of this thought. I would then be in a position to

[30] This recognition can perhaps be traced to Elizabeth Anscombe's paper "The First Person," in which Anscombe invented a mode of reference that was objectively but not subjectively reflexive. The paper is reprinted in Anscombe's *Metaphysics and the Philosophy of Mind: Collected Papers* (Oxford: Blackwell, 1981), 21–36.

[31] I think that this assumption is operative, for example, in John Campbell's discussion of "Self-Reference and Self-Knowledge," in *Past, Space, and Self* (Cambridge, MA: MIT Press, 1994), Chapter 4; and in Lucy F. O'Brien,"Anscombe and the Self-Reference Rule," *Analysis* 54 (1994): 277–81.

doubt whether "I" exist, since the doubt itself would guarantee only the existence of the doubter, who might not be the person whose existence was being doubted, however first-personally.

Fortunately, I needn't go so far as to suggest a gap between subjective and objective reflexivity. My point all along has been that secondary mental images have two subjects, one actual and one notional. The possibility of thoughts with notional as well as actual subjects requires us to enlarge our understanding of what it is for a thought to be reflexive.

The distinction between actual and notional subjects already figures in the subjective character of secondary images. Even to the imaginer himself, the image presents an imagining subject and a viewing subject, both in ways that are recognizably subject-presenting, and hence first-personal. So even within the category of subjective reflexivity, we must distinguish between actual and notional reflexivity, to mark the difference between the ways in which someone can be presented as the subject of thought.

We can then say that my mental image of Austerlitz, in its subjective character, is a notionally reflexive thought about Napoleon: I am thinking about NB in the notional first-person. And the notional first-person needn't refer to the actual subject of thought.

To claim that I can think of Napoleon in the notional first-person is still to claim too much, however. The notional reflexivity of my thoughts about Napoleon is less than genuine.

In order to imagine that I am Napoleon, I frame an image of Austerlitz as seen by someone who might thereby be moved to say "I see . . . ," and then I stipulate that the image and the associated utterance are oriented in such a way that "I" refers to NB. Without this referential stipulation, my mental image would not be a way of thinking about Napoleon as "me," and so it wouldn't be a way of imagining that I am Napoleon. Yet stipulations of this sort are foreign to reflexive usage. I don't usually specify to whom my uses of "me" refer – not even uses of the notional "me."

Suppose, for example, that I have a visual memory of a desolate field just like the one surveyed by Napoleon at Austerlitz. This memory includes a visual image that's presented as reprising an earlier visual experience, whose subject stood at the image's vantage point in front of the remembered scene. The memory image is thus presented as a duplicate, representing the field as seen by an original subject on some date in the past. It therefore has a notional subject, who would be the referent of the first-person pronoun in an accompanying image of the utterance "I see . . . ," if such an utterance were remembered from the same point of view.

If the image is indeed a copy of a visual impression, as it purports to be, then there is already a fact of the matter as to the identity of its notional subject: he is the person from whose experience the image was copied.[32] The image's notional reflexivity with respect to that person is not the product of any semantic stipulation on my part. I do not center the memory image on someone in the past so as to make him the notional subject. The image is just presented to me as having been copied from a visual impression, and it consequently represents things as seen by the subject of the impression from which it was, in fact, copied. Who he was is then determined by the image's causal history.

This mechanism makes the reflexivity of my memory genuine, I think, in a way that the reflexivity of imaginings is not. In memory I really think of the notional subject as "me"; in imagination, I only pretend to.

What makes a thought subjectively reflexive, after all, is that it is indexical in a special way: it has a peculiar way of pointing. A reflexive thought picks out a person at its center by mentally pointing to him in a distinctively inward-directed fashion. My experiential memories pick out past subjects by pointing to them in this way, but my imaginings cannot really do the same with Napoleon.

Before I can frame an image that points to Napoleon at its center – even its notional center – I must first frame another thought that picks him out, so that I can center the image on him. When I subsequently use that image to think of him at its center, I can only pretend to be using a mode of thought that's sufficient to pick him out. In fact, I couldn't have picked out NB as "me" without first picking him out as "Napoleon," in order to stipulate that he was the notional subject of thought.

Hence the thought of NB as "me" is less than genuinely reflexive. Genuinely reflexive thoughts don't rely on an antecedent specification of their target: they just point to the subject, at the center of thought. They are – to put it somewhat paradoxically – unselfconscious about their reference, in that they require no other thought about whom they refer to. I can think of NB as notionally "me" only by deliberately placing him where he will intercept this inward-directed pointer, thus rendering

[32] I do not mean to imply that the original viewer is the notional subject of the image solely because of its psychological origins in his experience. If the image wasn't presented in thought as the copy of a visual impression, then it might not present anyone as the notional subject, even if it was in fact copied from someone's experience. Because the image is presented as a copy, however, it has a notional subject, whose identity is then determined by his being the subject of the original. See also note 45.

its reference to him self-conscious. So I can only pretend to think of him in the notional first-person.[33]

4. Who I Was

But what if I believe that my memory is a vestige of Napoleon's experiences at the battle of Austerlitz rather than any experiences of my own?[34] In that case, I will believe it to be an image of Austerlitz as seen by Napoleon, on whom the image is centered naturally, without any stipulation on my part. And I will believe that it has a content that would be expressed by an accompanying memory of the utterance "I see Austerlitz," as spoken – and spoken truly – by a real person seeing Austerlitz. I may then transpose this utterance into indirect discourse by claiming to remember that I saw the battle of Austerlitz.[35]

[33] Note that the same considerations may apply to cases in which I imagine that I am David Velleman. For example, if I re-center my image of Austerlitz so as to imagine that I, David Velleman, am fighting in Napoleon's place, my thoughts do not become genuinely first-personal simply because they are now about DV rather than NB. I am still stipulating who is the notional "me," and hence only pretending to pick him out just by pointing.

[34] Gareth Evans argued that one could not question whether apparent memories derived from one's own experiences (*Varieties of Reference*, 235–48). According to Evans, one cannot even have a self-concept unless one is disposed to assimilate the information in memories and perceptions in ways that already constitute taking oneself as their source. A subject who didn't already treat himself as the source of memories, Evans argued, couldn't go on to doubt whether he was the source, since he would lack a concept needed for framing this doubt.

Note, however, that Evans's argument yields no conclusions about apparent memories taken singly. What the argument shows, if anything, is that I could not question whether I was the source of my recovered images in general. If I treat recovered images in general as derived from own experiences, however, then even by Evans's lights I will have the self-concept with which to doubt, about any particular image, whether I was its source. Hence Evans's argument does not preclude the possibility of my thinking that I have particular images recovered from Napoleon's experiences rather than my own. (Other potential obstacles to my taking this view are discussed in the following note.)

[35] Of course, I will also think that the image's content would be expressed by an accompanying memory of the utterance "I, Napoleon, see Austerlitz." Will I consequently claim to remember that I, *Napoleon*, saw Austerlitz?

Compare Andy Hamilton's remarks on the difficulty of reporting an apparent memory derived from Derek Parfit's experience of arriving at Bournemouth station:

> One could try 'I remember arriving at Bournemouth station – only the "I" then was Parfit!'. (It was the same 'I', only the person had changed his identity.) Or 'I remember arriving at Bournemouth, only it was not my body that arrived.' But these are desperate expedients. ["A New Look at Personal Identity," *The Philosophical Quarterly* 45: (1995), 332–49, 342.] [Note continues on p. 190.]

This report would be odd because the verb "to remember" is factive: the claim to remember something implies that it's true. If I speak the truth in claiming to remember that I saw Austerlitz, then what I claim to remember must be true as well; and what I claim to remember would seem to be that I saw Austerlitz. The merely bizarre belief that I have inherited one of NB's visual images seems to yield the truly absurd conclusion that I underwent one of his visual experiences.

One way to avoid such absurdities would be to qualify the description of my mental image. If I called it something other than a memory – say, an apparent memory or a quasi-memory[36] – then I wouldn't imply that it was veridical.

Yet my claim to remember that I saw Austerlitz wouldn't lead to absurd conclusions if it was properly understood. In saying "I remember that I saw Austerlitz," I am indeed claiming to occupy a mental state whose content is true. But I am not attributing to that state the content that would be conveyed by my saying "I saw Austerlitz" in *oratio recta*, where "I" would refer to the speaker, DV. Rather, I'm attributing to it the content that would be conveyed by an accompanying image of the utterance "I see Austerlitz," where "I" would refer to the original viewer. So I'm not reporting that I, DV, witnessed the battle of Austerlitz; I'm merely reporting memories of Austerlitz in which a witness of it is the notional "me."[37]

These are indeed desperate expedients, but only because they rely on an exchange of bodies or identities, which is quite unnecessary. What the subject of this transplanted memory should say is "I remember that I was Derek Parfit arriving at Bournemouth." This claim says nothing about an exchange between Parfit and the remembering subject, because – as I shall argue in the text – the second "I" is, not a reference to the rememberer, but a quasi-indicator echoing the first-personal conception under which Parfit's arrival at Bournemouth is being remembered. Similarly, my belief in having inherited Napoleon's visual image of Austerlitz should lead me to say, "I remember that I was Napoleon viewing the battle of Austerlitz." Again, the arguments required for a defense of this report are contained in the text, later.

[36] For the term "quasi-memory," see Sydney Shoemaker, "Persons and Their Pasts," *American Philosophical Quarterly* 7 (1970): 269–85. Actually, the mental states I am discussing would not be called quasi-memories by Shoemaker, because they are, as I put it, "recovered from" – and hence appropriately caused by – the original experiences.

[37] Thus, in "I remember that I saw Austerlitz," the second "I" is a quasi-indicator, which Castañeda would write with an asterisk, thus: "I remember that I* saw Austerlitz." So formulated, this statement begins to look like the formulations in Carol Rovane's "Branching Self-Consciousness" (*The Philosophical Review* 99 [1990]: 355–95, 368 ff.). According to Rovane, my image of Austerlitz would have to be reported as a quasi-memory of what "*I**" – rather than "I" – saw.

The resulting similarity between my view and Rovane's is potentially misleading, however. Rovane introduces "I*" as a "new pronoun" that is needed, she believes, because a

My mental image is indeed notionally reflexive with respect to such a person, if (as I believe) it was inherited from Napoleon. For in that case, the referential scheme of the image is not dependent on any prior specification of NB as the notional subject. Napoleon is the notional subject of my image because it is presented to me as derived from the visual experience of an original viewer, and that viewer was (so I believe) NB. His being the notional subject of the image is thus a matter of historical fact rather than stipulation; and so the image picks him out as "me" unselfconsciously, just by pointing to him in the center of its referential scheme.

Thus, if my mental image was inherited from Napoleon, then it represents Austerlitz as seen by a notional "me." I claim no more in saying "I remember that I saw Austerlitz." So why should I qualify my claim?

Some would answer that if I take myself to have an image of Austerlitz as it looked to Napoleon, then I shouldn't call it a memory, because a memory of how Austerlitz looked would have to be a memory of how it looked to me. In the view of these philosophers, experiential memory necessarily represents things as having been experienced by oneself, and it is "immune to error through misidentification" on this score.[38]

In my view, however, the nature of experiential memory can be fully explained by the fact that it represents things as experienced by a notional subject, whom it casts in the notional first-person, as "me." My memory of seeing something is necessarily a memory of *my* seeing it for the same reason that my image of being someone is necessarily an image of *my* being him – that is, simply because it is a first-personal way of thinking about the subject in question.[39]

report of what "I" experienced would pick out the subject of that experience as someone identical with me, the subject of memory. Since these subjects are not identical in this case, Rovane would have me replace the ordinary "I" with a different pronoun. In my view, however, the ordinary pronoun used in memory reports is the one that should be written as "I*," and it should be written this way precisely because it's a quasi-indicator that *doesn't* pick out the original subject as identical with me. I therefore deny that a new pronoun is needed: "I*" is just philosophical notation for the first-person pronoun as it is already used in memory reports. (For the same reasons, I shall also deny that there is any need for the notion of quasi-memory.)

[38] For these claims, see Shoemaker, "Persons and their Pasts," and Evans, *Varieties of Reference*, 235–48. More recent discussions include: John Campbell, "The Reductionist View of the Self"; and Andy Hamilton, "A New Look at Personal Identity."

[39] As P. F. Strawson put it: "[J]ust as nothing counts as an experience of a present state of consciousness which doesn't count as an experience of *being, oneself*, in that state of consciousness, so nothing counts ... as an apparent memory of a past state of consciousness which doesn't count as an apparent memory of *being, oneself*, in that state of

To be sure, such a memory cannot misidentify the viewer in representing him as me. But it cannot thereby misidentify the viewer, I say, only because it doesn't thereby identify him at all. A visual memory represents the viewer as me only in the sense that it represents the viewer *as the viewer*, who occupies first-person position in the visual scheme of reference. The original viewer was "me" in this sense no matter who he was, just by virtue of being the notional subject of the image; and his having been "me" in this sense does not entail his having been DV. Memory can thus succeed in making someone "me" to me even if he was Napoleon – not, of course, by making him the same person as me, but rather by presenting him to me in the notional first-person.

The assertion that experiential memory can make Napoleon "me" to me sounds like Locke's assertion that memory makes a person "self to himself" across time. It therefore suggests a way of re-interpreting Locke's theory of personal identity, by suggesting a perspectival sense in which one can be "self to oneself."[40]

The word "self" has two related but ultimately distinct strands of meaning. It connotes both identity and reflexivity, and either of these connotations might dominate when the word serves as a noun. On the one hand, a past self of mine might be one and the same person as me, identified at some time in the past. On the other hand, a past self might be someone in the past whom I can think of reflexively, in the first-person. In the first sense, selfhood is a metaphysical relation that holds between persons at times, if they are the same person. In the second sense, selfhood

consciousness.... What we have here is an enriched version of Kant's repeated point about the 'I think' merely being the form of consciousness in general" ("Kant's Paralogisms: Self-Consciousness and the 'Outside Observer'," in *Theorie der Subjectivität*, ed. Konrad Cramer et al. [Frankfurt:Suhrkamp, 1987], 203–19, 216–17).

[40] *An Essay Concerning Human Understanding*, ed. Peter H. Nidditch (Oxford: Oxford University Press, 1975), Book II, Chapter xxvii. See also the following passage, in which Kant criticizes the notion that first-personal thought reveals the existence of a persisting mental substance:

> Despite the logical identity of the 'I', such a change may have occurred in it as does not allow of the retention of its identity, and yet we may ascribe to it the same-sounding 'I', which in every different state, even in one involving change of the [thinking] subject, might still retain the thought of the preceding subject and so hand it over to the subsequent subject. (*Critique of Pure Reason*, trans. Norman Kemp Smith [New York: St. Martin's Press, 1965], 342.)

This passage is related to Locke's argument purporting to show "that two thinking Substances may make but one Person" at different times (*Essay*, 338). As Kant's version of the argument makes clear, however, what the argument really shows is that different thinking substances could be accessible to one another's first-personal thought – which, as I am about to suggest, makes them one and the same *self*.

is a psychological relation that holds between subjects who are on first-personal terms.

Memory really does make a person "self to himself" in the latter sense. When I entertain experiential memories, I have thoughts that present a past individual to me in the notional first-person. Memory thereby recruits past selves for me, by putting them within reach of subjectively reflexive thought.

Locke's memory theory is thus a correct account of perspectival selfhood. Of course, Locke clearly intended the theory to be a metaphysics of persons. But what if he confused the two?[41] Maybe Locke got perspectival selfhood right but then mistook it for personal identity.[42]

In order to minimize confusion, let me divide the available meanings between the terms "selfhood" and "personal identity." From now on, I'll use "selfhood" to denote the relation borne to me by those whom I can think of first-personally – my grammatical person-mates, so to speak, whom I shall call "selves." I'll use "personal identity" for the relation among those who are one and the same person, and I'll describe them as the same person rather than as selves.

If Locke had been clearer-headed, he might have offered a theory of selfhood and left it at that. This theory would have had nothing to say about whether Napoleon and I are the same person; but it would have had plenty to say about whether Napoleon was among my past selves. Napoleon was a past self of mine, the theory would have said, if I have memories derived from his experiences and can therefore think of him in the first-person, just by pointing to him unselfconsciously as "me."

Of course, Napoleon wasn't really a past self of mine. My memory of surveying a desolate field may make me think that he was, by making me think that he is the referent of the first-person in its referential scheme.

[41] This interpretation of Locke was suggested by Elizabeth Anscombe in "The First Person," 25–26. The present essay can in fact be read as an attempt to salvage something of interest from the confusion that Anscombe identified in Locke. For a different theory of selfhood as based on reflexivity rather than identity, see Robert Nozick, *Philosophical Explanations* (Cambridge, MA: Harvard University Press, 1981), 71–114.

[42] Here I do not mean to imply that Locke's metaphysics of persons is necessarily wrong. Indeed, one might argue that Locke ended up getting the metaphysics of persons right by thinking in perspectival terms. For under some conceptions of what persons are, their persistence through time might reasonably be thought to depend on relations of first-person accessibility between temporally disparate points of view, and hence on perspectival selfhood. Yet to say that persons are entities whose identity depends on perspectival selfhood is to make a substantive philosophical claim, which must not be obscured by a conflation of the metaphysical and perspectival notions. (In fact, however, I do not think that a theory of perspectival selfhood can serve as a theory of metaphysical identity without some modification, for reasons that are explained in note 53.)

But the referent of "me" in my memory image is the subject from whom the image has been inherited, and that person wasn't really NB.

In reality, let's suppose, my memory is derived from a visual experience received on Breed's Hill in 1976, during a 4th of July celebration reenacting a Revolutionary battle. The battlefield represented in my memory image must therefore be Breed's Hill rather than Austerlitz, and the referent of "me" in the image is the person who stood at its vantage point, undergoing the visual experience from which it is derived – DV, as it happens, rather than NB.[43]

Since NB is not the person whose encounter with the depicted scene produced this image, he is not the notional subject of the image, and the image doesn't recruit him as one of my former selves. He can of course be an imaginary self of mine, since I can pretend to have notionally reflexive thoughts about him. But these thoughts would not be genuinely reflexive with respect to NB, because they would have to be self-consciously centered on him before they could point to him, at their center, as "me." Because I am not really on first-personal terms with Napoleon, he is not really one of my former selves.

A clearer-headed Locke might have offered this theory of selfhood, but would we have had any use for it? Isn't personal identity what we really care about? If so, the Lockean theory of selfhood would have been true but pointless.

I now want to argue that this theory would not have been pointless, because selfhood is of independent philosophical interest. Indeed, I think that some of the deepest concerns expressed in terms of personal identity are actually perspectival concerns about the self.

In order to address these concerns, however, Locke would have had to extend his theory slightly. For they are primarily concerns, not about whose past we are remembering, but rather about whose future, if any, we are in a position to anticipate. And addressing these concerns would have required Locke to extend his theory from the past selves who are recruited by memory to the future selves who are recruited by anticipation.

5. Who I Will Be

What we most want to know about our survival, I believe, is how much of the future we are in a position to anticipate experiencing. We peer up

[43] On this point, see Hidé Ishiguro, "Imagination II," *Proceedings of the Aristotelian Society* Supp. Vol. 41 (1967): 37–56, 43, 52.

the stream of consciousness, so to speak, and wonder how far up there is still a stream to see.

To wonder how much of the future I can anticipate experiencing is just to wonder how far into the future there will be experiences that I am now in a position to prefigure first-personally. If this question truly expresses what I want to know about my survival, then what I want to know is a matter of perspective rather than metaphysics. My question is not how long there will be an individual identical with my present self, DV. My question is how long there will be someone to occupy the position that is the center of my self-centered projections – someone to serve as the referent of "me" as it occurs in my prospective thoughts. The future "me" whose existence matters here is picked out precisely by his owning a point of view into which I am attempting to project my representations of the future, just as a past "me" can be picked out by his having owned the point of view from which I have recovered representations of the past.

One complication is that in the context of anticipation, the reference of "me" may not be determined as it is in the context of memory. "I" refers to the notional subject in either case, but the notional subject may not be determined in quite the same way.

Suppose that while preparing for this year's 4th of July celebration, I anticipate my role in the annual reenactment of a Revolutionary battle. I conjure up a mental image of the climactic moment – the field, the tunic, and so on. In its intrinsic features, this mental image is no different from that in a memory or an imagining. What differentiates it from these images must be how it is presented.[44] Whereas the image in a memory is presented as the vestige of a past experience, for example, the image in anticipation must be presented – or intentionally framed – as prefiguring a future experience.

In the case of memory, we noted, the presentation of an image does not fully determine its references. Even when I think that I'm recalling Napoleon's experiences at Austerlitz, my memory is not an image of Austerlitz if it is actually derived from a glimpse of Breed's Hill.[45]

[44] On the question of how an image is "presented," see note 18.

[45] Of course, what places the references of an image under the control of its causal history may be its presentation as a recovered experience. After all, an image that was actually derived from a glimpse of Breed's Hill could subsequently be incorporated into an imagining of Austerlitz – in which case, its causal history would not prevent the imaginer's intention from making it refer to Austerlitz instead of Breed's Hill. But when an image is presented as reprising a past experience, its references are thereby hitched to its origins in experience, despite concomitant misjudgments as to what those origins might be. (Here I am indebted to Michaelis Michael for his objections to a purely causal analysis of a memory's references.)

But the reverse appears to be true of images framed in anticipation. My anticipatory image is of the forthcoming military maneuvers precisely because I think of it as prefiguring my experience of those maneuvers. The presentation of this image may even consist in an intention on my part, which places the image's references under my voluntary control. For I may conjure up the image with the express intention of thereby prefiguring the experience of playing my role in the reenacted battle – in which case, the image is of playing my role, as I intend.[46]

In this respect, anticipation appears to resemble imagination, whose references are similarly determined by an accompanying intention or stipulation. Unfortunately, this resemblance seems to prevent anticipation from providing a context in which I can think about future individuals unselfconsciously as "me." In framing a mental image with the intention of prefiguring a future experience, I have to specify the experience to be prefigured. And in order to specify the experience, don't I have to specify its subject?

If so, I will end up deliberately centering my image on someone, and then it won't be a genuinely first-personal thought about him, since I won't have picked him out simply by pointing to him at its center. He will be at most an imaginary self of mine. Perhaps, then, my future selves are all imaginary.

I think that there are indeed modes of anticipation in which I project myself into the perspective of the future DV in a manner no different from that in which I can project myself into just anyone's point of view. In these cases, anticipating my future amounts to no more than imagining the future life of DV. But there are other modes of anticipation, I think, which are quite unselfconscious about the future perspectives that they prefigure, and which consequently place me on genuinely first-personal terms with future subjects. I shall argue that these modes of anticipation ground a distinction between real and imaginary future selves.

One such mode of anticipation is that in which I frame an intention to do something in the future. Framing an intention entails projecting myself into a future perspective because it entails representing the intended action from the point of view of the agent who is to perform it.

Of course, the agent who is to perform any action that I intend must be me, since I can't intend the actions of others. But intentions of doing

[46] I may therefore enjoy infallibility with respect to the references of my anticipation. See Wittgenstein's remarks on this subject in *The Blue and Brown Books* (Oxford: Blackwell, 1972), 39.

something are always intentions of *my* doing it, I would argue, in the same sense as memories of seeing something are always memories of *my* seeing it – namely, in the sense that these attitudes always have a notional subject, whom they present as "me."

Intentions always have a notional subject because their function is to be acted on, and they can be acted on only if they are drawn from the agent's point of view. Intentions are consequently framed in a referential scheme centered on their potential executor, who is thereby thought of as "me," no matter who he will be.[47]

Intention resembles memory, furthermore, in that I do not have to stipulate who its notional subject shall be. For if my intention is going to be executed, its executor will have to be the person who finds himself in possession of the intention when the time for executing it arrives.

An intention must be framed on the assumption that it or its mental traces will persist until they can serve as a basis for action.[48] In framing an intention, then, I project my thoughts into the future in two distinct senses. On the one hand, I project my thoughts into the future in the sense that I represent the world from a specified future point of view. On the other hand, I project my thoughts into the future in the sense that I *send* them into the future, by depositing them in memory for future retrieval. And the point of view into which I mean to project my thoughts in the first sense is simply that point of view into which I shall have projected them in the second. That is, I mean to represent an action from that perspective at which this representation will, at the relevant moment, be availaƚ 'e as a basis on which to act.[49]

[47] This statement oversimplifies a very complicated story. In many cases, intentions cannot be framed from the executor's perspective, because his perspective cannot yet be fully envisioned. For example, I may intend to go north in the future because I cannot yet envision whether going north, at the relevant point in my travels, will entail going left or right, backwards or straight ahead. But if I intend to go north, my intention is incomplete, precisely because it will have to be translated into self-centered terms before I can act on it.

[48] This assumption need not be distinct from the intention, since part of what is intended may be precisely that this very intention persist until it can be put into action. See, for example, Gilbert Harman's view that intentions refer to themselves as causes of the intended actions (*Change in View: Principles of Reasoning* [Cambridge, MA: Bradford Books, 1986], 85 ff.).

[49] To speak of the perspective at which the representation itself will be available is of course to presuppose a theory of diachronic identity for mental representations – which may be too much of a presupposition in this context. But my references to the storage and retrieval of a single, persisting representation can be replaced with references to a momentary representation and its causal descendants at later times. The language of persisting representations is just an expository convenience.

Thus, I don't have to specify a person from whose point of view I am trying to frame my intention, because that point of view is fixed by the future causal history of the intention itself. I attempt to frame the intention, if you will, from the intention's own future perspective, the perspective in which the intention itself will turn up to be executed. Just as a memory purports to represent the past from the perspective at which it originated in experience, so an intention purports to represent the future from the perspective at which it will arrive to guide action. In either case, the relevant perspective is picked out by the natural history of the representation itself; and the referent of "me" in the context is simply whoever fills the role of subject within that perspective.

As it happens, of course, the perspective at which any intention of mine will turn up to be executed, and from which I have therefore tried to frame it, will belong to the future David Velleman. This older DV will turn out to occupy the position of notional subject in my intention, and so he will turn out to be the person of whom I was thinking first-personally in the context. Being accessible to unselfconscious first-personal thought on my part, he qualifies as my real future self.

The double projection that characterizes intentions is not confined to practical thought, however. Even when I am just picturing the future, without planning to do anything in it, I usually regard my mental image as entering into a future perspective both representationally and causally. I don't just anticipate experiencing the future; I anticipate experiencing it as the payoff of this anticipation, as the cadence resolving the present, anticipatory phrase of thought. Now, a musical phrase is resolved by its final notes only for a listener who is still mindful of how it began. So when I anticipate experiencing the future as resolving this anticipation, I picture it as experienced from a perspective in which this picture is recalled.

This mode of projective thought has a look and feel all its own. Within the frame of my anticipatory image, I glimpse a state of mind that will include a memory of its having been glimpsed through this frame – as if the image were a window through which to climb into the prefigured experience.[50] Anticipating the future in this manner, I once again look to future selves unselfconsciously. I don't specify the notional subject of my anticipatory image. He is simply the person who will confront the envisioned future with this image at his back, glimpsed in memory as the

[50] This "window" is unfortunately not a WYSIWYG environment: What You See looking through it Is not necessarily What You Get upon climbing through.

image through which his state was glimpsed in anticipation. And he is a real future self of mine because, as the one who will experience the imagined future from the other side of this image, he is picked out by the natural history of the image, as the person whom it presents in the notional first-person.

Finally, my allusions to future subjects can be unselfconscious without necessarily involving the thought that they themselves will be remembered. My prior image of an event may produce various other thoughts, emotions, or inclinations whose remnants will color a future experience of the event even if no memory of the image itself remains. I can then picture the event as experienced in the psychological wake of this picture, whether or not a memory of itself will be among the items that the picture leaves in its wake.

If the wake of an experiential image is expected to wash over the prefigured experience, the image may then be constrained in what it can justifiably portray. I'm hardly entitled to anticipate an event as being experienced with shock and disbelief from a perspective that will have been influenced even indirectly by this anticipation, since the event is unlikely to incite either shock or disbelief in a mind bearing the traces of its having been hereby anticipated. Conversely, there may be events that I'm entitled to anticipate as being met with equanimity only from a future perspective that will retain traces of this anticipation.

What will transpire in perspectives that intercommunicate with mine in this fashion matters more to me than what will transpire in other perspectives. Indeed, my epistemic relation to these perspectives may partly constitute their mattering to me. To imagine a future pain, for example, as it will feel in the psychological wake of my hereby imagining it is to do more than just imagine it. It's to imagine the pain as befalling a mind that has somehow been prepared by this very prospect of its occurrence. And to imagine a pain as experienced by a mind hereby so prepared for it is already to brace for the pain, to shrink from it, or to be otherwise caught up in it in some way. Anticipation that's cognizant of its effect on the prefigured experiences is thus a form of mental engagement with them that, to some degree, already constitutes their mattering.

This engagement with future experiences coincides, of course, with an ability to regard their subjects unselfconsciously as "me." When I frame an image prefiguring an experience that will follow in the image's wake, causally speaking, I needn't specify for whom the experience will follow: in the context of the image, the experience is simply "to follow" – to

follow the image itself, that is. The image thus prefigures the experience simply as forthcoming, and so it provides a context for thinking about the subject of that experience unselfconsciously as "me."

6. Why "Me"?

In sum, anticipation that engages its object tends to be genuinely first-personal, and vice versa. This association may help to explain why I care about my future selves: they are the persons whose experiences I cannot prefigure without already being caught up in them, as lying in the wake of this anticipation.

But the association between selfhood and engaged anticipation is merely an association, which can sometimes fail, if not in reality, then at least in imaginary circumstances. The question therefore arises whether I care about my selves only in virtue of my psychological engagement with them. Or do I care about my selves as such?

The best way to approach this question will be to entertain an imaginary case in which selfhood and psychological engagement come apart. I will therefore conclude with a brief discussion of a familiar philosophical fiction.

Imagine that my brain will be divided and each half transplanted into a different body, with the result that two people will wake up tomorrow remembering my past and carrying on my anticipations and intentions for the future.[51] If I know what is in store for me, I can frame anticipations today that will have effects on, and perhaps be remembered in, two different perspectives tomorrow. Hence I can actively anticipate the future as experienced by two different people.

Even so, I cannot make either person the notional subject of my anticipations unselfconsciously. Suppose that I try to think ahead into some future moment at which I shall have two psychological successors. If I try to picture the moment as it will appear in an experience specified merely as forthcoming, or to follow, I won't succeed in picking out the perspective from which I'm trying to picture it, since my picture may be followed, in the relevant sense, by two different experiences of the moment in question, and I cannot be trying to draw it from both perspectives at once. Similarly, my anticipation may be remembered in two

[51] See David Wiggins, *Identity and Spatio-Temporal Continuity* (Oxford: Blackwell, 1967), 50; and Parfit, *Reasons and Persons*, 254 ff. Parfit says (fn. 40), "I decided to study philosophy almost entirely because I was enthralled by Wiggins's imagined case."

different perspectives, and so I cannot frame it from a perspective specified merely as that in which it will be remembered.

In order to specify the perspective from which I'm trying to picture the future, I'll have to identify it with one of my psychological successors or the other.[52] That is, I'll have to pick out the person whose perspective is the intended target and destination of my projective thoughts – whereupon I'll be doing exactly what I do when imagining that I am Napoleon. My anticipation of the future will be nothing more than an act of imagination.

By depriving me of unique future perspectives, fission would deprive me of real future selves.[53] It wouldn't prevent me from being fully engaged with both successors, however, since both lie in the causal path of my present thoughts. The question is whether anticipatory engagement with them would preserve all that matters about survival. Would I suffer a significant loss in having no subject with whom I was on genuinely first-personal terms?

My inclination is to say that I would indeed suffer a loss. I could no longer think just about how the future would look; I'd have to think about how it would look to particular, specified observers. I could no longer plan just to act; I'd have to plan actions to be performed by particular, specified agents. I could no longer imagine a future as existing simply on the other side of this image; I'd have to imagine it as existing on one or another of the image's "other sides," in the lives of one or another of my psychological successors.

Here I am tempted to borrow again from Bernard Williams, by saying that my relations with successors-by-fission would always involve "one thought too many." Williams coined this phrase to express the loss of intimacy that a Kantian moral agent would suffer in relations with others.[54]

[52] This point figures prominently in Rovane's "Branching Self-Consciousness."

[53] Note that first-person reference is asymmetrical in this case. Although I cannot refer first-personally to the products of my fission, they can refer first-personally to me, in the context of their experiential memories. This result strikes me as intuitively correct. When I imagine undergoing fission tomorrow, I don't seem to have much of a future; but when I imagine that I am the product of fission that occurred yesterday, I still seem to have a complete past. (This intuition is shared by Simon Blackburn, "Has Kant Refuted Parfit?") This result also demonstrates that selfhood, defined perspectivally, cannot coincide with the identity of a person, since selfhood turns out to be asymmetric whereas relations of identity cannot.

For the claim that "creatures involved in fission and fusion could have nothing like our ordinary use of the first person," see John Campbell, *Past, Space, and Self*, 97. Campbell bases this claim on very different grounds.

[54] "Persons, Character, and Morality," in *Moral Luck* (Cambridge: Cambridge University Press, 1981), 1–19.

I, too, am using the phrase to express a loss of intimacy, but the intimacy lost in this case would be in relation to my own psychological successors, and the excess thought would simply be the thought of who they were. In cases of fission, I would have to identify particular successors before I could enter their perspectives: there would be no future perspectives that I could enter without a second thought. And the second thought of whose perspective I was entering would be an alienating thought, one too many for the intimacy that holds among selves.

In some respects, of course, I would still be in a position to anticipate the lives of my successors "from the inside," as we sometimes say. In particular, I would be able to project my thoughts into their perspectives both causally and representationally, sending into their points of view images drawn from those points of view. But in another respect I would no longer be in a position to anticipate any future life from the inside, since there would be no life that I could anticipate without first picking it out for the purpose of projecting myself into it. Surely, a position from which I must deliberately project myself into a life is not a position on the inside of that life.

My sense, then, is that the ability to prefigure future experiences unself-consciously is an important part of having a future at all. Not being just plain "me" to myself would be more than the loss of a pronoun; it would be the loss of a self-intimacy that is part of what matters about having future selves.