Velleman: Self to Self

Critical Study for *Nous*, 44:4 (2010):1-19

John Perry

§1. INTRODUCTION

The title essay of David Velleman's *Self to Self* (Velleman, 2005) deals with thoughts and concerns about oneself. These issues are importantly related to Velleman's famous account of practical reasoning, the subject of much of the rest of the book, for he takes gaining self-knowledge to be a primary motivation of the most deeply human kinds of such reasoning. I will focus on the ideas he develops in the title essay, "Self to Self". I'll explain Velleman's theory, then offer various criticisms and suggestions.

§2. VELLEMAN ON SELF TO SELF

Velleman's aim is to provide an account of our self-regarding concern about the future. What am I concerned about, if I am concerned whether I will survive another few months to see Obama inaugurated? Velleman distinguishes two answers:

i) Whether the person I now regard as self (i.e., think of as *me*) will survive until January 20, 2009;

ii) Whether there will be a future person, (i.e., a person existing on January 20, 2009) whom I can now regard as self.

The first answer, which I'll call the "straightforward answer," seems to require two bits of apparatus, the concept of a person at a time regarding a person at that very same time as self, and then the concept of survival or personal identity. The second answer, which is Velleman's, requires the concept of a person at a time regarding a person at a future time as self --- an instance of the "self-to-self"

relation. We need to understand what Velleman takes this relation to be, and why he thinks it is crucial to our self-regarding concern about the future.

Velleman talks about past and future selves; I'll assume to talk about a self at a time is just a way of talking about a person existing and having experiences at that time; no special ontology of momentary selves seems required. That is, the talk of selves allows us to identify persons by the roles they play in distinct episodes of consciousness, and so raise questions of identity in an intuitive way.

Velleman distinguishes, among thoughts that a person X has about X, between those that are (as I like to put it) about "the person X happens to be," and those that are genuine self-thoughts, those X would naturally express with the first person. If Elwood finds a coded memo from his dean to his chairman that says Philosopher Z is going to get a big raise, and wonders if Philosopher Z really deserves a big raise, and as it happens he is Philosopher Z, then he has a thought that is about himself in the first way. If he reads on, and from various details realizes that he is Philosopher Z, he will think about himself in the second way; if Elwood is like most of us, his doubts about desert will be resolved positively, and he'll be happy.

The fascination with the difference between thoughts about the person one happens to be and genuine self- thoughts goes back (at least) to Hector-neri Castañeda, and Velleman relies on Castañeda's concept of "quasi-indication" (Castañeda, 1966, 1967). Suppose I say, "Elwood believes that he is going to get a big raise." This seems false, or at least misleading, with respect to the first part of the story in the last paragraph, but definitely correct with regard to the second. Castañeda's diagnosis was that we ought to distinguish two words, "he" and "he*". The first is the ordinary pronoun, which here has "Elwood" as antecedent and refers to him, but does not tell us anything a about how Elwood is thinking about the person he is thinking about, only that it happens to be Elwood himself. If that's the "he" I am using, my remark is true about the first and second parts of the story. "He*" requires, in addition, that Elwood is having a genuine self-thought. If I was using "he*" then my report was false about the first part of the

story, but true about the second. "He*" is what Castañeda calls a "quasi-indicator".

We can recognize the phenomenon of quasi-indication, and two uses of "he," and even use "he*" to clarify our reportorial intentions, without taking a position on the vexed issue of whether there are two pronouns, or one ambiguous pronoun, or simply one unambiguous pronoun, with quasi-indication being a matter of pragmatics rather than semantics; on the latter view my remark is literally true, but in most cases misleading, with respect to the first part of the story. For Velleman's purposes, as he rightly sees, it is the phenomena, and not the semantic and pragmatic issues, that are important.

In both uses, the "he" is reflexive; that is, the person thought about is the thinker him*self*. Velleman distinguishes between objective and subjective reflexivity; in subjective reflexivity, the thinker is genuinely self-thinking about himself or herself.

What makes a thought a case of genuine self-thinking? Velleman holds that genuine self-thinking involves thinking of oneself as the subject of one's mental images, images of the sort involved in ordinary perception, imagination, memory, and anticipations of various sorts, including intention. Elwood has an image of the memo he has found; he sees it in a certain way, from a certain perspective. As he puts out his hand to turn the page, he sees the hand in a certain way, and is also aware of the movement of the hand in a certain way. Elwood thinks of himself quite spontaneously as the subject of these images; as the person who occupies the center of the perspective they supply. In the first part of the story he isn't thinking of Professor Z as the subject of these images, but in the latter part of the story he is. He is thinking of Professor Z as himself.

Elwood's perceptual images are what Velleman calls "primary images". In the case of memory, anticipation and imagination *secondary* images are also involved. Secondary images have both an actual and a *notional* subject. Suppose as Elwood reads on, he is flooded with memories of reading notices from his bank about bounced checks, seeing the faces of his hungry children, and the like.

Elwood is actually having the memories, so he --- his present self --- is the actual subject. But some past self actually had the remembered images, that is, was the actual subject of the perceptions of which they are memories; that self is the notional subject.

Among the notional subjects, there are those that are genuine and those that are stipulated. The notional subjects involved in ordinary memories of what one has seen and done, or anticipates seeing and doing, are genuine. But suppose, as part of imagining being Napoleon, Elwood has images of the battle of Austerlitz, of the sort he supposed Napoleon to have. Elwood *decides* who he is imagining being; perhaps he would have had the same images, had he imagined being Napoleon's lieutenant. Napoleon is the *stipulated* notional subject.

A genuine self-to-self relation obtains between Elwood and selves that are the genuine notional subjects of his images; that is, the selves whose experiences he remembers and anticipates in the normal ways we remember and anticipate our own experiences. So when Elwood remembers seeing the faces of his hungry children, he has a self to self relation with one of his own past selves. But when he imagines being Napoleon watching the battle of Austerlitz unfold, he does not have a genuine self to self relation with Napoleon.

We can, Velleman says, distinguish two relations between Elwood the note-reader and Elwood the children-perceiver. The children-perceiver is the same person as Elwood, identified at the past time. That is a metaphysical relation. Elwood the note-reader also has a self to self relation with Elwood the children-perceiver; the latter is a self the former can think of reflexively, with genuine self-thoughts. This is a psychological relation, that holds between subjects who are "on first-personal terms". Memory makes a person "self to himself" in this psychological way; the past individual is presented to the rememberer in the "notional first person."

Intentions of the ordinary sort, (and anticipations of the future more generally), always have notional subjects; that is, the intention is framed in terms of the perspective of the potential executor. "Just as memory purports to

represent the past from the perspective from which it originated in experience, so an intention purports to represent the future from the perspective at which it will arrive to guide action." (198)

In the case of intention and anticipation, there is no need for Elwood to single out the person whose experiences he anticipated or whose actions he forms intentions about; the matter is not stipulated; it is unconscious. Being "accessible to unselfconscious first-personal thought," qualifies the future self as being Elwood's "real future self. "(198)

In both memory and intention there is a double relation to the future self; there is the psychological self-to-self relation and there is in addition a causal relation; the past thoughts, actions, and perceptions of the earlier Elwood's are causally responsible for his current memories; his current intentions will be causally responsible for his future actions.

Typically, but not necessarily, the anticipations involve anticipations of memories:

> This mode of projective thought has a look and feel all of its own. Within the frame of my anticipatory image, I glimpse a state of mind that will include a memory of it having been glimpsed through this frame --- as if the image were a window through which to climb into the prefigured experience. 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 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. (198-199).

All of this, Velleman maintains, is a central part of the explanation of 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."

§3 TWO INSIGHTS

It seems to me that two valid insights form the basis of Velleman's theory. One is that the straightforward account needs some supplementation. We need to understand what it is to think of a certain person as *self*. It is natural to suppose that such an account will emphasize the most primitive way we have of thinking about ourselves, as the subjects of the perceptions we are having, and the agent of the actions we are performing. But the role of being the subject of our present perceptions and agent of our present actions is not the same as the role of being the subject or agent of past and future actions. The straightforward account needs to provide a concept of self-thinking that can be extended to the past and the future.

Further, this extension isn't just a matter of knowing who we are, and thinking of *that person* as having perceived and acted in the past, or as perceiving and acting in the future. To see this, suppose I have become deluded, and take myself to be John Searle. Then John Searle is, in a clear enough sense, the person I think of as *self*. Even so, there is a difference between my hope that John Searle will live to see Obama inaugurated, and my hope that *I* will live that long.[1] In fact I, the non-deluded version, have both hopes. And if the deluded me were freed of the delusion, he would continue to have both hopes. And finally, the deluded me could well have the hope that John Perry will live to see Obama's inauguration, even he himself does not. So the hope that *I* will live to see Obama inaugurated is not simply the hope that *John Perry* will live to see Obama's inauguration. Even if I know I am John Perry, and so will naturally hope that *I* live long enough, if I hope John Perry does, the two are separable. It is the hope that *I* will live that long, that is crucial in understanding the special kind of hope that counts as the hope of survival. So the straightforward account cannot solve

the extension problem by, so to speak, routing our past and future regarding thoughts through the persons we happen to be.

The second insight is that self to self relations play an important part in the phenomenology of human thought, that play and important role in understanding the special concerns we have for our own pasts and futures.

I think, however, that Velleman illicitly combined these insights, and tried to use the self to self relations as a general account of the extension of self-thinking into the past and the future, and I do not think this works.

§3. THE PHENOMENOLOGY OF SELF-THOUGHT

I think Velleman's account will not quite do as a *general* account of self-thought with regard to the present, past, and future, for a couple of reasons. First, as a general account, I think it gets the phenomenology wrong. Self-to-self relations, as described by Velleman, are not always, or even typically, involved in memory and intention; or, even if they can be somehow superimposed on all such episodes, they seem to get at important underlying structure only in relatively special cases. Remembering past experiences, in the first-personal way, does not always, or even usually, involve reliving the past experiences in a way that straightforwardly provides a notional subjects, nor are such notional subjects typically involved in a straightforward way in intention.

With regard to intention, Velleman says, "Intentions are consequently framed in a referential scheme centered on their potential executor, who is thereby though of as "me," no matter who he will be." In a footnote to this passage, Velleman says that it is an oversimplification, because often the executor's perspective cannot be fully envisioned. Suppose Elwood forms the intention to thank the dean, for his wisdom in recognizing Elwood's value, once the raise has been fully put in place. Elwood has no idea how the opportunity for gratitude will present itself. An accidental face-to-face meeting in the quad? An encounter in the dean's office after some committee meeting? A phone call? A letter? An email?

In these cases, typical of intentions to act in the even the slightly remote future, the concept of the notional subject of the experiences at the time of the execution doesn't seem to get much of a grip.[2] Elwood is not in a position to contemplate the future experiences, the perceptions and the like, that will be occurring at the time of execution, except by thinking of them as what he* will be experiencing whenever the opportunity for fulfilling the intention presents itself. Which of Velleman's options seems to fit Elwood thoughts better as he wonders about the prospect for getting the dean thanked?

i) Whether the person I now regard as self (i.e., think of as *me*) will at some point in the future thank the dean;

ii) Whether there will be a future person who thanks the dean, whom I can now regard as self.

It seems that ii) doesn't fit Elwood's case, if we take the requirement "whom I can now regard as self" in the way Velleman explains it. This would involve thinking about the dean-thanking episode in a certain way, which defines a perspective and a notional subject, which one will unconsciously and automatically take to be oneself. Elwood will automatically and unconsciously take the dean-thanker to be himself, but this doesn't seem to be because he thinks of the future episode of dean thanking in a rich enough way to define a perspective; he seems to lack the materials to do so.

In the footnote mentioned, Velleman notes that in so far as Elwood doesn't know the circumstances in which the opportunity for gratitude will present itself, his intention, or at least his plan, is incomplete, "precisely because it will have to be translated into self-centered terms before I can act on it". The point may be that Velleman's account of intentions it is only meant to apply to complete intentions. If Elwood is to carry out his intention, it will be by making precise movements at the time; walking across the quad, or approaching the dean after a meeting, or making a call, dialing a certain number, or writing a letter or an email to the right address. By the time he executes the intention, Elwood will have constructed a complete plan, much of it at the last minute on the fly, when he sees

the dean across the quad, or realizes that now would be a good time to call him. Does Velleman's account apply to such complete intentions?

Imagine a case where the formation of Elwood's intention coincides with grasping the circumstances in which it will be carried out, so he is in a position to form a complete intention. Suppose Elwood sees the dean one day across the quad, and just at that moment decides to thank him. The plan will build back from the intended action, thanking the dean, to actions that will, in Elwood's circumstances, constitute a way of thanking the dean. Elwood sees the dean a ways off, decides to thank him, recognizes that being near the dean and saying "Thanks for the raise; I was wondering when we would have a dean wise enough to recognize my merits," is a way of thanking the dean; recognizes that changing his path across the quad in a way to intercept the dean is a way of putting himself in a position to carry out his intention, and begins walking.

The terms "actual subject" and "notional subject" are both *role* terms; that is, they identify a person by the role they play with regard to a situation or episode. When Velleman thinks of past or future selves, he thinks of them as conscious, experiencing persons, who are playing the role of subjects with respect to some or all of the mental phenomena involved in the separate episodes. In the case of Elwood on the quad, it is artificial to divide the flow of mental activities into two episodes, two selves, the intender and the executor. Elwood of course plays many roles relative to this episode. He is the possessor of the intention, the perceiver of the dean, the possessor of the sensations and anticipations involved in that perception, the filler-in-of the plan, the walker, the executor of the intention, the extender of the hand, and the like. On Velleman's analysis, things start with an episode of Elwood having an intention and looking forward to its execution; the role of actual subject is filled by Elwood as intender, while Elwood as executor of the anticipated execution is the notional subject. At the end there is an episode of executing the intention while remembering forming the intention, with Elwood the executor as actual subject, and Elwood the intender as notional subject. These concepts can doubtless be superimposed on the episode, but I

don't see that they disclose any very essential structure; that is, analyzing Elwood's thought in terms of two episodes and four roles seems like a bit of overkill.

What definitely is correct about Velleman's analysis is that self-thought involves unconscious role binding. Let me say a little bit more about roles and role-binding, before elaborating on this. Think of driving your car on the freeway. You see some red lights and congestion on the hill about a half-mile away; you notice the gap between your car and those in front narrowing, you step on the brakes, and prepare to turn off on a shoulder if necessary.

If we made your practical reasoning completely explicit; that is, detailed all the contingencies on which the success of your plan depended, we would have to separate in analysis many roles you don't consciously separate in thought. Consider the facts that a certain car is the car in which you are seated, the car controlled by the brakes attached to the pedal your foot is on, the car whose position you can ascertain by looking through the windshield in front of you, and the car whose front wheels are controlled by the steering wheel in your grip. Those are all contingencies; they are the result of a very rational design for the ways cars are put together; a body of design insights that even Detroit has not yet managed to screw up. You are attuned to all of these facts; that is, you take actions that depend on them to further your goals; they are built into the way you expect the world to work in response to your actions; but this is unconscious, you don't need to think about them. If some day you become the remote controller of a dozen or so cars, whose positions are disclosed to you by a bunch of computer screens, so you are required to push buttons or adjust pointers to determine which car you are affecting, you will have to unlearn these habits.

Much of our practical life is based on unconscious attunement to multiple roles we can count on a single object to satisfy. The computer I'm looking at is the one that is effected by moving my fingers on the keyboard beneath them; the object I see in my hands is the one I feel with my hands; the place whose weather I see out the window is the place whose weather I will experience when I go out

the door. The prankster (who plugs my keyboard into your computer) and the artificial intelligencer (who has to write thousands of lines of code to get the roles properly linked in a virtual reality world) need to be aware of the contingent nature of these role-bindings, and the philosopher should be too, so as not to miss the epistemological richness below the surface of our conscious thought.

Philosophers have sometimes tried to give simple role-based analyses of our sense of self. After all, a role-based analysis of the word "I" seems to work well enough: the referent of "I" is relative to an utterance is the speaker of that utterance. But in fact we play many roles relative to episodes of thought, perception and action, some, perhaps, bound together of necessity, many as a contingent matter. I am the person seeing *these* things; the person having *these* sensations; the person performing *these* actions; the person experiencing *these* emotions; perhaps, as Russell emphasized, the person having *these* sense-data and and Broad did, the person having *this* inchoate background of bodily feeling, and much else besides. Daniel Dennett, a conceptual prankster, artificial intelligencer, and philosopher, showed us how we lose our sense of self when these roles get distributed and separated in odd ways in "Where Am I?" (Dennett, 1978a).

When we form intentions, or just anticipate the future, the way we think about ourselves involves the binding of many roles. When I reach for the glass, a part of the success conditions of my movement is that the person seeing the glass, the person having the thirst, and the person whose hands can be directly caused to move are all one and the same. The role-bindings are unconscious of course, although philosophy or virtual reality experiments may bring them to consciousness.

4. PLANNING BACKWARDS

Velleman's account of intention seems most apt for those cases in which we begin by vividly imagining what it would be like to, say, jump from an airplane, or speak Russian, or be hooded as part of a graduation ceremony. We start by imagining these things in some detail; the adrenalin rush, or the sense of

accomplishment and feelings of pride, the smile on parents' faces, the looks of surprise or relief on our advisors' faces. In such cases, the vividness of imagining the future experience can motivate us; we want to be a person, or the person, who has such experiences. (Or perhaps we vividly imagine what it will be like to die of lung cancer, or awake at the end of a cold night spent in a drunken stupor in a back alley, and resolve not to ever have those experiences.) We start not simply with goals we want to accomplish, but the future experiences achieving those goals will involve, and build a causal line back from them to our present situation, and so put ourselves in what Velleman calls a self-to-self relation. We build a bridge between the notional subject of the imagined futures and the actual subject of our current intention-forming frame of mind. It seems to me that ordinary planning doesn't work this way; we start from the here and now, and plan forward, usually in a rather piecemeal fashion, focusing on the goal rather than--- at least in any detail---the experiences that we expect to have at the time we achieve it. So I'll call the special case "backwards planning."

It may be that for some people, the concern to live on is tied to these sorts of anticipations and backwards planning. I want to live to see the Second Coming, or at least last day of the Bush Presidency, or my youngest grandchild's graduation from college. So, in this sense, Velleman's concept of self-to-self relations does, as he claims, show us a something important about why and how we are concerned to survive. But it seems to me it does not show the basic structure of ordinary intentions, or the universal basis of the concern we have for our future selves.

§5. FIRST PERSON MEMORY AND MEMORY FROM THE INSIDE.

Similar points apply to memory. Our typical locution for intention is, "X intends to A" where it is implied that the A-er, in case X's intention is fulfilled, will be X himself or herself. It seems logically there should be a quasi-indicator: "X intends that he* A". And some linguists would say that there is such a thing in "deep structure;" the subject of the action has been "deleted under identity".

In the case of memory, we have the locution "X remembers A-ing", as in "Obama remembers growing up in Hawaii". This works similarly to the standard locution for intention; logically, one might suppose, it should be something like "Obama remembers his* growing up in Hawaii." In any case, this memory locution, which I'll call "first-person memory," is part of a more flexible family than "intends." We can also say that Obama remembers that he grew up in Hawaii," and that Obama's grandmother remembers Obama's growing up in Hawaii. So, limiting ourselves simply to episodic sorts of things, we have first-person memory (X remembers A-ing), memory-that (X remembers that Y A-ed) and event-memory (X remembers Y's A-ing). These different locutions impose different conditions. If Obama remembers growing up in Hawaii, then he must have grown up in Hawaii. If Elwood remembers Obama growing up in Hawaii, then he must witnessed Obama growing up in Hawaii (or at least been aware of it at the time). If Wynona remembers that Obama grew up in Hawaii, then it must be true that he did, and she must have learned this earlier; but she doesn't need to have grown up in Hawaii or ever witnessed anyone doing so.

First person memory, as I use the term, gets at a way of reporting memory, suitable for reporting a person's memories of her own past thoughts, experiences, and actions. I'll use Sydney Shoemaker's phrase "memory from the inside" to get at a certain phenomenology in which the past events we were involved in, may present themselves. Memory from the inside occurs when we have present images, that are, or at least purport to be, copies of the sensations, perceptions, emotions and other experiences we had at the time; in other words, we at least seem to remember from the very perspective we had at the time of the remembered event.

Now it seems to me that remembering from the inside in this sense is not logically required for first person memory, and, in fact, though not rare, is a relatively unusual way to remember our events that are even slightly remote. When I remember graduating from high school, I visualize the event from a perspective no one occupied, somewhere over Pinewood Bowl in Lincoln,

Nebraska; there is a line of young people in black gowns; there is a very tall broad-shouldered boy --- that's Henry Pangborn --- and behind him a shorter, skinner boy --- that's me. In this case memory from the inside would be exceptionally boring; at the time I could see nothing but the black gown draped over Henry's broad shoulders. Such memory, when it occurs, is largely reconstructive. That is, it isn't as if the sensations and experiences I was having were stored away in memory, and then just pop up again, so that I relive the event. I use the information I carried away about the event, and what I know about how things work, to reconstruct the experiences I must have had: seeing the back of Henry's gown, hearing crickets (how could there not have been crickets on a June evening in Nebraska), smelling the Pine trees (it was Pinewood Bowl, after all) and the like. It's not that the particular look and feel and smell, and more commonly the emotions, involved in an experience, are never retained; if Henry had fallen backwards and crushed me, I'd probably retain not just the fact that it happened but some trace of the terror and pain of the event.[3] Reconstruction is still likely to be involved in filling in the details.

So, I don't see much room for the self-to-self relation playing an essential role in the general case of first-person memories, any more than in the case of intention. In the case of memory, unlike that of intention, the self's experiences involved in the episode are fully determinate before the present thinking, but by and large detailed, perspective-determining images are not involved. But of course they can be, and perhaps the cases in which they are, are of particular importance.

§6. LOCKE AND VELLEMAN

John Locke is usually credited with formulating the problem of personal identity, and advocating a "memory theory" as a solution (Locke, 1694). However, Locke doesn't use the phrase "remember," but rather the phrase "extend our consciousness backwards." He subscribes to the principle that we never have a thought or experience without being aware that we have it. That awareness of our own present thinking and experience is the consciousness that gets extended

backwards. Was he thinking about the general phenomenon of first person memory, or more specifically about memory from the inside? The language suggests the latter. When one thinks of "short-term memory," it seems like we do have memory from the inside, and indeed seems like we must for memory to do its job. As Elwood walks to intercept the dean, for example, his actions at each moment are suited to information about the location of the dean relative to himself that demand more information than might be retained in long term memory. A day or a week after the event he may remember that he thanked the dean on the quad, but may well not remember, at least not without some reconstructive work, where he was on the quad, and where the dean was on the quad relative to him. But as he walks towards the dean, perhaps taking his eyes off of him while he straightens his tie, or walks through a crowd of students, he is guided by information that accumulates from the past and is augmented by his new perceptions.

The situation is somewhat similar to that we noted in discussing intention, and perhaps we should distinguish long term intention from short term intention. In the case of both memory and intention, the information we access is suited to the use we need to put it to. In memory, the information necessary to walk in the right direction, at the right speed to intercept the dean, is necessary at the time of the episode, but pretty irrelevant later. That evening Elwood will want to tell his wife that he met the dean on the quad and thanked him, but unless there was something particularly striking about the details of the episode, he won't have much use for recreating anything very close to the perspective he had at the time. Intention is similar, with the time reversed.

Locke's intentions are not so clear. Did he think that to incorporate a remotely past thought, experience or action into one's sense of self, one needed to remember it from the inside, or simply to remember it in the way ordinarily reported with first-person memory? As the theory Locke suggested has developed, philosophers have tried to capture logically necessary and sufficient conditions for personal identity in terms of memory. Those who sympathize with

"Locke's memory theory," usually seem to have first person memory in mind. But Locke's own interest seems to have been largely in explaining how the merit and blame for past thought and action can be incorporated into one's later self-concept, and perhaps this means that some of the steps philosophers have used to develop his theory would not appeal to him. Reid challenged his theory with the brave officer paradox; an officer remembers stealing apples as he bravely picks up the fallen standard on the battlefield; years later, as a retired general, he remembers picking up the standard, but no longer remembers stealing the apples. So, Reid says, the general seems to both be the child who stole the apples (since he is the officer, and the officer was the child), and not be that child (since he can't remember stealing the apples). Subsequent Lockeans have suggested we take something like the ancestral of "remembers" or "can remember" as the condition for personal identity, thus evading the paradox.[4]

But would Locke approve? This weakened condition, he might think, doesn't do the work for assigning responsibility he had in mind --- the forensic role of personal identity, as he put it. Insofar as Locke was exploring the phenomenology of responsibility and feeling of merit and guilt for past actions, he might find Velleman's concept of self to self relations more pertinent to his concerns than the constructions of subsequent Lockeans. Velleman suggests that Locke would have been well advised to develop the phenomenological side of his theory without suggesting that he was offering a solution for the metaphysical problem of personal identity. This is an interesting and valuable insight.

§7. IMAGINING BEING NAPOLEON

Velleman explains many of his key concepts about self to self relations in memory and intention by reference to, and by contrast with, the case of imagining being someone else, in particular, the case of imagining being Napoleon at Austerlitz. The question with which he wrestles is this: when DV imagines being Napoleon at Austerlitz, where does DV fit into the content of the imagining? His conclusion is that he doesn't fit anywhere. The images mind of the battlefield and the fallen soldiers and brave deeds and horses dashing about belong to DV; they are his

imaginings. But what he is imagining is Napoleon's experience; DV isn't a part of what DV is imagining. It's not a real self to self relation, because the images he has aren't really Napoleon's; Napoleon is the imagined subject, and Austerlitz the imagined battlefield, as a result of the stipulation that is involved in the imagining. Napoleon is not a genuine notional subject of the imagining.

The main motive for this approach is the fact that, aside from very strange speculations about reincarnation or something like that, it really isn't a possibility that DV be Napoleon. So there are no possible worlds in which DV is Napoleon. If we think of the content of an attitude of imagining as something like a proposition that is defined by the set of possibilities in which what is imagined is true, then DV is imagining nothing. But it doesn't seem that he is imagining nothing. So we need to find an alternative content.

I think that once one realizes the generality of the problem involved here, one will think that Velleman is indeed onto something, but his account needs to be supplemented. Suppose, for example, Elwood is landing in Buffalo, but takes himself to be landing in Cleveland. While he was asleep, the pilot announced that the plane had been diverted, because of bad weather in Cleveland. Elwood looks out the plane and thinks and says, "That city is Cleveland". Straightforward semantics takes "that city" to refer to the city he sees and demonstrates, that is, Buffalo, "Cleveland" to refer to Cleveland, and the proposition expressed to be the absurdity that Buffalo is Cleveland. I'll call this the "official" content, since the semantics involved is widely accepted and due to authorities like Donnellan, Kaplan and Kripke. But the official content doesn't seem to capture the relevant content of Elwood's thought. For one thing, his thought seems to rationally motivate certain actions that Elwood takes; he turns to the page of the airline magazine that has a diagram of the Cleveland airport; he gets his Cleveland guidebook out of his backpack; he resolves to call his sister in Cleveland, who was going to pick him up, as soon as the plane lands and cell phones may be turned on. Intuitively, these actions and intentions are rational, given his desires, in that if his belief is true, they will further those desires.

Intuitively, it seems that these are actions that make sense the way Elwood takes the world to be; that is, in terms of possible worlds, in the worlds that fit his beliefs, his actions promote his goals.

The official content, the impossibility that Buffalo is Cleveland, doesn't seem to explain the rationality of Elwood's actions; it seems no more relevant to them than any other necessary falsehood, which provides no possibilities that, if actual, would make his action fruitful --- say that George W. Bush is General Grant, or that DV is Napoleon.

What seems to be needed here is what we might think of as back-up content. The problem we are dealing with is more or less the dual of the one that bothered Frege. That problem was how A=B can be more informative than A=A, when A=B is true. The present problem is how A=B can be misinformative when it is false; that is, misinformative in ways that rationalize some actions and not others. Frege's idea was that there is an alternative content, involving a mode of presentation associated with "B" that was not associated with "A"; the proposition that a single object is presented twice is the informative, back-up content. If we set aside the details of Frege's brilliant but somewhat controversial development of this idea, in his theory of *Sinn* and *Bedeutungen*, his idea seems clearly on the right track.

In Elwood's case, when he thinks, "that city," he is thinking of a certain city, a city that happens to be Buffalo, as the city he now perceives through the window of the airliner. His referential plan, one might say, is to refer to Cleveland, by referring to the city he sees through the window; the plan fails because it is based on a false premise, that the city he sees is Cleveland. If this premise were true, then the city at which he is about to land would also be Cleveland, and learning about Cleveland's airport, studying the guidebook, and calling his sister all make good sense. Elwood is thinking of, and referring to a certain city, Buffalo, as it happens, in a certain way, and that way of thinking provides us with our backup content.

Elwood is in the window seat. The passenger in the middle seat might say to the one in the aisle seat, "That man thinks that city is Cleveland". The phenomenon of quasi-indication is involved here. In this situation, the reporter's use of "that city" is a quasi-indicator, that not only refers to Buffalo, but gets at how Elwood is thinking of Buffalo --- whether semantically or pragmatically is again, not important for our purposes. The point is that the aisle-sitter would understand what is going on: Elwood has a belief that would be true if the city he were seeing out the window were Cleveland --- a perfectly coherent possibility, that explains why Elwood is poring over the diagram of the Cleveland airport. And, when he says to Elwood, "that city is Buffalo," Elwood won't be likely to reply, "Well I already knew that Buffalo is Buffalo." He will understand what they intent to convey to him: that the city he is seeing through the window isn't Cleveland, but Buffalo.

If we are "direct referentialists" of some sort or another, we will want to recognize the official content of Elwood's remark; we will think that when Elwood says "that city is Cleveland" the necessarily false proposition that Cleveland is Buffalo deserves some special status. But there is no reason to suppose that that is the only available content.[5] Any assessment of the content, or truth-conditions, of an utterance depends on what we hold fixed, and what we allow to vary. Given the "that city" refers to Buffalo, and "Cleveland" to Cleveland, Elwood's utterance is true only if Buffalo is Cleveland, which it cannot be. But if we allow the reference to vary, Elwood utterance is true if whatever city he is looking and attending to is Cleveland, which could have been so. Anyone who picks up the quasi-indicative message of "that man thinks that city is Cleveland," is grasping this back-up content. They would realize that Elwood's actions make sense on the supposition that the city he is looking and attending to is Cleveland.

Now suppose Elwood heard the pilot, and realizes that they are landing in Buffalo, but wishes that they were landing in Cleveland. He might express this wish, intelligibly enough, by saying "I wish that city were Cleveland (and not

Buffalo)." He might even find himself imagining that it was Cleveland. In what sense does Buffalo enter into what is imagined, that is, into this alternative, back up content? Only indirectly, as the actual referent of the thought, "that city". But importantly. What Elwood imagines is not so, and it is not so because Buffalo is the actual referent of "that city," and Buffalo, for better or worse, is not Cleveland. The official content and the back up content are both false, the first necessarily so, the second contingently.

Let's return to my delusion of being John Searle. I admire John Searle, and have for a long time. At times I imagine being John Searle (or at least, as he seems to me): confident, assertive, with many important books to his credit, as at home in Paris as in Berkeley, with impeccable judgment in most things, and confidence in all. Perhaps at some point in the future, my imagining will give way to delusion, and I'll think that I am John Searle. For years I strode into the Philosophy Lounge with a look of confidence, imagining that I was Searle; but eventually I will stride in truly confident, thinking I am Searle. The content of my earlier imagining seems like it is the content of my later delusion, the false proposition that I express with "I am John Searle." The more or less official content of this utterance, given our understanding of the first person as a tool of direct reference in more or less Kaplan's sense[6], is the impossible proposition that John Perry is John Searle, but that doesn't get at the important content of my imaginings or my delusion.

When I am deluded, I plan on referring to John Searle by using the word "I" in its ordinary sense, that is, by referring to the speaker, to myself, to the person that has these thoughts, sees these things from this perspective, and the like. I think that by referring to myself I can refer to John Searle. It is this alternative, back-up content that explains why, given my delusion, my behavior in defending Searle's views on mind and body, and my sharing with graduate students what (seem to be) memories of it was like to be at Oxford in the days of Austin, makes sense.

Here I am the actual referent of my use of "I", and I do intend to refer to myself. But I intend to thereby refer to Searle, and in that I don't succeed. It is commonplace in the philosophy of action to distinguish a number of things that the agent does, or tries to do, by taking various circumstances as fixed, or allowing them to vary, depending on what we are trying to understand. We distinguish between the actual result of an act, and the intended result. What a person does, the result of his actions given the actual circumstances, may appear unintelligible, given his beliefs and desires; it becomes intelligible when we consider what would have happened had the circumstances been as he thought they were. When I say, "I really enjoyed writing *Intentionality*," I try to say something true, and would have, if the speaker of my very utterance had been Searle; but instead I refer to John Perry, and say something false. The philosophy of language needs to take flexibility lessons from the philosophy of action.

When DV imagines being Napoleon, watching the events of Austerlitz unfold, what he imagines involves DV in the same way that his delusion would, were he deluded, actually thinking that he was Napoleon. What he does imagine, and what he would believe, are false. Velleman is the actual referent of the thoughts he expresses with "I". As Velleman observes, when he is imagining, he is imagining referring to Napoleon by referring to himself. When he is deluded, he thinks he can refer to Napoleon by referring to himself.

When he imagines being Napoleon at Austerlitz, what DV imagines is false, and necessarily so, at the level of official content, and DV is a constituent of that content. But DV is not a constituent of the explanatory, back-up content, and this is what seems to be to be correct about Velleman's analysis. The back-up content is roughly that the subject of *this* episode of thought (the one DV is conscious of, in the way Locke thought we were conscious of all of our present thoughts and actions), is Napoleon, and he is surveying the battlefield at Austerlitz. DV is not a constituent of this content. The constituents are an episode of thinking and imagining, and Napoleon. If what is imagined, at the back-up level, were true, then when DV says, "I am winning," or "I am

Napoleon," what he said would be true. DV is nevertheless importantly involved, as the person who is actually playing the role of subject of the episode.

When we say, "DV imagines that he* is Napoleon at Austerlitz" the "he*" actually refers to DV. The statement provides us with two false contents for DV's imagining. The official content is necessarily false; there is no (relevant) possible world in which DV, the imaginer, is Napoleon, the victor. The back-up content, the one we use to understand the episode of imagining, the one the quasi-indicator provides by giving us the relevant role, is contingently false; there are worlds in which the roles of being the subject of DV's present thoughts and being Napoleon the victor at Austerlitz are filled by the same person, but the actual world isn't one of them.[7] So DV has four roles to play; he is the imaginer; he is thus the referent of the reflexive quasi-inicator "he*", thus he is a constituent of the official content of the sentence embedded in the that-clause, and, although he is not a constituent of the back-up content, he is the actual player of the role involved in the back up content, being the subject of the episode of imagining.

The moral of these reflections on imagination is this. The supplementation of the straightforward accounts needs, to deal with past and future regarding thoughts involving oneself, won't be found at the level of the official content of the thoughts involved, but at the level of backup content.

§8 THE SELF AND THE FUTURE

I'll end by trying to develop of version of the straightforward account that can incorporate Velleman's insights, and the considerations that emerged in our discussion of imagination.

Let's start with concept employed in the first part of Velleman's option i), the person I now regard as self, or think of as *me*. Earlier we distinguished between two ways a person X can think about X, thinking of the person he happens to be, and genuine self-thinking. I think we need to make a further distinction, between primitive self-thinking and ordinary self-thinking.

Primitive self-thinking is thinking about the world from the perspective one has on it, and does not require a concept or as I prefer *notion* of oneself. An animal sees the world from a certain perspective, and thus gains information about how things are in relation to the animal that occupies the perspective. But this doesn't require the animal to have a separate concept of itself. A fairly sophisticated animal may recognize various objects, and accumulate information about those objects, and so form expectations of what interactions with those objects will yield, based on previous interactions. Such an animal needs to distinguish among similar objects and keep track of their differences. To that extent, he needs notions of the objects he recognizes, accumulates information about, and acts differentially towards in light of that information. But all the information he gathers through perception provides information about himself; he doesn't need a notion of himself to keep track of who is it is that he food dish or the tree he sees is in front of, he doesn't need to recognize which agent he is getting information about.

Adult humans pick up information in more complex ways that require a self-notion to keep track of things. We each have a concept of ourself, as one among the many people there are, the one who occupies the perspective that our perceptions give us information from, and we accumulate information about ourselves in a dedicated self-notion. If I see my name in a phone book, or on a poster for a lecture, or in a time schedule, I can pick up information about myself in the same way I pick up information about others, by finding a name, and examining the sentences and other entries in which it occurs. I can get knowledge in this way about the person named ``John Perry'' --- about his phone number, or the time of his lecture or his class--- and he is the person I happen to be. But normally, unlike the Castañeda cases, I recognize that the information is about me. This means that I integrate the information (or, often enough, misinformation) into my self-notion, the repository of perceptual information, and the information I get through memory and by forming intentions. Self-knowledge, for adult humans, ordinarily involves such a self-notion.

Even so, primitive self-knowledge never goes away. We don't need to recognize ourselves as the perceivers of the object we perceive, or the initiators of the actions we initiate, any more than the animal does. Nor do we normally need to keep track of whom it is our first person memories are about, or who the executor of the intentions we form will be.

The thought that I will see Obama inaugurated, and the thought that John Perry will, have the same referential or official content. They differ in back-up content. What makes my thought that *I* will see the inauguration about me is the fact that it involves my self-notion. What makes my self-notion about me is simply that is mine. It may be quite inaccurate in many ways, as in the case where I have incorporated a lot about John Searle into it. But it is mine; when I think using it, my thoughts are about me, just as when I use the word "I", my assertions are about me, however deluded I may be, and whoever I may be trying to refer to with "I".

My thought that John Perry will live to see the inauguration is also about me. But it is about me because I have a notion of John Perry, as one of many people that have wandered about the earth, and John Perry is the source of that notion. Since I know who I am, my self-notion and my John Perry notion are linked, but they could become separated, as I acquire amnesia, or sink into delusions of being Searle. It is the backup contents of my two thoughts that differentiate them, and it is at the backup level we can understand the different roles they play, or might play, in my thought and action.

So, to conclude. Velleman's concept of self to self relations seems to me to illicitly combine two important but separate insights. One is that the straightforward account is insufficient, unless we incorporate into it the special way we ordinarily think of ourselves, via our self-notions, which are tied to the primitive ways we have of gathering and using information we have about ourselves; a way of knowing that is tied to our role in our own perceptions and actions, so that the subject of the present perceptions and the agent of present actions does not need to be recognized, and the subject of past perceptions and

future actions does not need to be *stipulated*; the various roles are linked, as Velleman says, unconsciously. The second is that there is a way of thinking of past and future thought and action, the self to self way, in which we think of ourselves as the occupant of the perspective that we had or expect ourselves to have, that is phenomenologically important and important to the special concerns we have about our pasts and futures. What seems to me to be mistaken is taking this special attitude, the self to self phenomenon, as the key to replacing or supplementing the straightforward account.[8]

REFERENCES

Bratman, Michael. (1999) *Faces of Intention*. Cambridge: Cambridge University Press.

Castañeda, Hector-Neri. (1966) "'He': The Logic of Self-Consciousness," *Ratio*, vol. VIII, No. 2: 130-57.

Castañeda, Hector-Neri. (1967) Indicators and Quasi-Indicators. *American Philosophical Quarterly*, 4 (1967): 85—100

Dennett, Daniel. (1978a) Where Am I. In Dennett (1978): 310-323.

Dennett, Daniel (1978b). *Brainstorms: Philosophical Essays on Mind and Psychology.* Montgomery, VT: Bradford Books.

Locke, John (1694). Of Identity and Diversity. Chapter 27 of *Essay Concerning Human Understanding*, Second Edition. Reprinted in Perry (2008): 33-52.

Perry, John (2008). (Editor) *Personal Identity,* Second Edition. Berkeley, University of California Press.

Perry, John (2001). *Reference and Reflexivity*. Stanford, Ca.: CSLI Publications.

Reid, Thomas (1785). Of Mr. Locke's Account of Our Personal Identity. Chapter 6 of "Of Memory," which is the third essay in *Essays on the Intellectual Powers of Man.* Reprinted in (Perry, 2008): 114-118.

Shoemaker, Sydney (1970). Persons and Their Pasts. *American Philosophical Quarterly* 7 (4): 269-285. Reprinted in Shoemaker (1984): 19-48

Shoemaker, Sydney (1984). *Identity, Cause and Mind.* Cambridge: Cambridge University Press.

Velleman, David. (2005) *Self to Self.* Cambridge: Cambridge University Press.

---

[1] Written in 2008.

[2] This point is emphasized in Michael Bratman's theory of intentions as plans (Bratman, 1999).

[3] One may, under hypnosis, or in psychoanalysis, have quite vivid memories that present themselves as relivings of the past experiences. Perhaps this shows that many memories are in fact stored away in phenomenological detail, which can be accessed with effort. It may only show that in these circumstances our reconstructions may present themselves vividly. The memory-like experiences produced by these methods are not always accurate.

[4] See the essays by Quinton, Grice, Shoemaker and Parfit and Perry in (Perry, 2008).

[5] See (Perry, 2001).

[6] I say "more or less" because for his purposes in developing a logic, Kaplan emphasizes that his is a theory of sentences in contexts (in his sense), not a theory of utterances, but I am thinking of utterances.

[7] This depends on its making sense that episodes of thought and consciousness could have different subjects than they actually do. This is clearly epistemically possible, as cases like the Searle delusion show. One might argue that it is not metaphysically possible, on the grounds that episodes of consciousness are *individuated* by their subjects. I believe that (virtually) all arguments in philosophy that employ the word "individuate" in any substantial way are fallacious, but I will not try to make that case here.

[8] I am grateful to the participants of the seminar Michael Bratman and I gave on Velleman's book at Stanford in Spring 2006 for help in formulating the ideas developed in this study, especially to Bratman and to Sarah Paul.