6

Perry on Self-Knowledge

NEIL VAN LEEUWEN

1 Problems of Self-Representation

A man is looking out the large window at planes on the tarmac. The PA system announces, "John Perry to the counter at Gate 27A." The sound waves strike the man's eardrums and are translated into neural impulses. His brain signals his leg muscles, which contract and carry him to the counter. He speaks, "*I* am John Perry."

The man must believe that *he* is John Perry. This is a clear case of self-belief, which is self-knowledge when it is true, justified, and whatever else is needed for knowledge.[1]

But what is self-belief? The intuitive answer is: Self-belief is just belief about an individual, where that individual is the same person as the person who has the belief. But this answer is insufficient.

To see the insufficiency, let's fictionally alter our initial example. *a-John*, as I'll call him, is an amnesiac version of John Perry. a-John is sitting in the airport and has forgotten he is John Perry. Hearing the announcement, a-John forms the belief that *John Perry should go to the counter at Gate 27A*. Yet a-John remains sitting. Due to amnesia, he doesn't know or believe that he is John Perry.

Could we rectify the situation just by giving him more information about who John Perry is? Say someone, attempting to help, stopped and told

[1] This is a broader sense of self-knowledge than the sense in which self-knowledge refers only to knowledge of one's own mental states. I explain the relation between those senses in the next section.

The Signifying Self: An Introduction to the Philosophy of John Perry.
Albert Newen and Raphael van Riel (eds.).

him that John Perry is a philosopher with a white beard, who happens to be sitting near 27A. a-John could then go on to infer *the philosopher with a white beard near 27A should go to the counter*. But he still doesn't stand up, because he doesn't realize he is that philosopher.

We can now see why the intuitive characterization of self-belief is insufficient. a-John has a belief about an individual—that that individual should go to the counter—and the individual that belief is about is the same as the individual who has the belief (a-John/John Perry). Yet this still isn't self-belief, which is why he is unmoved.

Our problem is thus clear. We need to sufficiently characterize self-belief.

a-John does of course have self-beliefs. He believes he is sitting down; he believes he is in an airport; he believes he is wearing shoes; he believes he feels warm; he believes he'd like to smoke a pipe. So we need to say how these genuine self-beliefs differ from the beliefs that, although they're about him, are not self-beliefs.

a-John has three notions in his mind that are *of* him, which I notate[2] as follows:

[John Perry]
[white-bearded philosopher near 27A]
{self-notion}

Each of these notions figures as a constituent in beliefs a-John has, but it would seem the true self-beliefs employ the self-notion. For example:

{self-notion} am [in an airport]

So we can say: Self-belief is just belief about an individual, where that individual is the same person as the person who has the belief, *and* the individual is represented in the belief by the self-notion.

Have we solved our problem? Not yet, for even though the claim just made is true, an analogous problem arises at the level of notions, which are ways of thinking about something. The intuitive characterization of self-notions is this: A self-notion is a notion about an individual, where the individual who has the notion and the individual it is about are the same. But this is insufficient, since a-John's notions [John Perry] and [white-bearded philosopher near 27A] both satisfy the intuitive characterization, without being the self-notion. When a-John thinks *John Perry*, he's thinking about himself without knowing it, so the [John Perry] notion is *not* the self-notion.

---

[2] I'll clarify the difference in meaning between the square and curly brackets shortly.

We haven't solved the initial problem of self-belief, but we've sharpened it. We now need to give a theory of what a self-notion is. The plan for the rest of the paper is to use John Perry's (the real-life philosopher's!) ideas to do just that.

2 The Plan: Introducing Perry on Self-Knowledge

A central theme in Perry's philosophy is that there are often many different ways to think (or talk or write) about the same thing. Understanding the different ways systematically is critical to understanding cognition and communication. Developing this theme as it applies to thinking about the self, I structure this paper around three questions.

> *First*, how should we classify the different ways of having information about the self? There are several ways I can receive and retain information about me. What are the ways?
> *Second*, how does self-knowledge fit in with the structure of knowledge, belief, and cognition, in general?
> *Third*, how can we characterize the self-notion, using the general picture of knowledge, belief and information retention at our disposal?

Answering the second question is crucial to answering the third. To see why, let's return to our example of a-John, whom we'll leave behind entirely after this section. By analyzing how a-John can come to realize who he is, we can start to see what it is for a notion to be a *self*-notion.

There a-John sits, accumulating information. Some information he accumulates is tied to the name "John Perry". a-John looks in the bag next to him and sees a book on the philosophy of John Perry, with a picture of a man with a white beard on it. He learns Perry is a philosopher of language with a white beard.

But other information he acquires is immediately cognized as being about himself. Touching his chin, he realizes that *he* has a beard. Seeing his reflection in the window, he learns his beard is white.[3] Looking around he sees *he* is in an airport at gate 27A. Thinking rather deep philosophical thoughts about language, he realizes that *he* is a philosopher of language.

Suddenly, a light bulb goes on. He exclaims, "I am John Perry!"

What just happened?

---

[3] Of course, seeing one's reflection only feeds information into the self-notion file if one realizes it is a reflection, as *a-John* does in this case. We'll see Perry's treatment of the Ernst Mach example below, in which Mach didn't realize he was seeing a reflection. In that case, the reflection-generated percepts (initially) don't feed into the self-notion file.

There are (at least) two kinds of information channel active in a-John's pre-realization state. a-John's realization occurs when the two channels become connected in the right way.

Channel 1. On the one hand, a-John sees big signs with symbols like

27A

which are encoded in perceptual states. He also feels a seat underneath him and sees the airplanes outside the window. This inflow of information is from his immediate environment and his own body and relates that environment and body to *him*. For example, his 27A percept represents the sign as being a certain distance away, and although he's not thinking of himself in having that percept, *he* is the entity from whom the 27A sign appears as being that certain distance away. So—this is the critical point—the 27A percept (along with a host of others like it) is guaranteed to convey information that is at least in part about him: It locates him in an environment. Let's call this kind of information *agent-relative*, since the informational content always relates the agent receiving the information to a surrounding environment. The agent receiving the information is *identical* with the agent *to whom* the inflowing information relates the environment. Even amnesiacs can use the agent-relative information flow to form self-beliefs, with reflexive contents like *I am in an airport*. (Note to the reader: I use curly brackets "{…}" to designate notions in the mind that derive from this agent-relative information channel.)

Channel 2. On the other hand, some of the information a-John receives is not relative to him at all—at least not until he finds a way to connect it to himself. He might read that interest rates have gone up in Singapore, that on right triangles $a^2 + b^2 = c^2$, or (crucially) that John Perry is a philosopher. This kind of information we can call *objective* (or *detached*, for reasons that become clear), since the information gleaned from this flow isn't without further realization taken to relate to the agent him- or herself.[4] (Note to the reader: I use square brackets "[…]" to designate notions in the mind that are associated with objective information.)[5]

---

[4] It happens to be the case that objective information flow is always carried by a channel of agent-relative information flow. For example, learning the objective information *that there is coffee in Colombia* requires being in a perceptual state: either of seeing the assembly of letters "there is coffee in Colombia", or of hearing that sentence spoken, or of being in Colombia and seeing coffee beans. But assume for now that we can separate out the objective portion of the channel.

[5] This use of the phrases "agent-relative", "objective", and "detached" comes from Perry's own work; but the use of the curly and square brackets to designate notions associated with either agent-relative or objective information is my own convention.

When a-John realizes *he* is John Perry, he must be noticing that there is simply too much coincidence between the information from the objective flow about John Perry and the information from the agent-relative flow about himself and his place in his environment. Both he and Perry have white beards, think philosophical thoughts, have something to do with gate 27A, etc.

When a-John realizes *I am John Perry*, he is thinking something like:

> {the individual seeing 27A, feeling the bench, hearing the PA …} am [the philosopher named John Perry, a man with a white beard, a man who must come to 27A …]

Otherwise put:

> {self-notion} am [John Perry notion]

What this suggests is that the self-notion is constitutively linked to agent-relative information—in a way that has yet to be explained. Furthermore, a-John has his realization *by* linking the objective information about John Perry *to* the self-notion and agent-relative information. In short, the suggestion is, roughly, that {the individual seeing 27A, feeling the bench beneath, hearing the PA announcement …} *is* a-John's self-notion. As Perry writes: "[S]elf-notions are those that have the special role of being the repository for information gained in normally self-informative ways and the motivator for actions done in normally self-effecting ways" (Perry 2002c: 205). I call this *Perry's Thesis*.

It will take work to flesh out this thesis, but the discussion so far already gives an idea of Perry's methodology. First (in order of logical priority), he distinguishes different kinds of information channel or "ways of knowing", including ways of knowing that for "architectural" reasons give information about the person using them; second, Perry also distinguishes different classes of action, including what he calls "normally self-effecting ways of acting"; third, he uses the ways of knowing and categories of action to implicitly define the self-notion, as standing in certain relations to those ways and categories (this strategy is essentially functionalist); fourth, self-belief (and hence self-knowledge) can be explained by appeal to the self-notion.

3 Question 1: What Are the Different Ways of Having Information About the Self?

The phrase "self-knowledge" often refers specifically to knowledge of one's own mental states. But Perry's use of the phrase is broader and, in a sense, more fundamental. Perry generally uses "self-knowledge" to mean any

knowledge one has of one's self, where that knowledge has the self-notion as a constituent: knowledge one might express using "I" or "me". So my knowing I'm wearing a blue shirt is self-knowledge in this extended, Perryan sense. Importantly, however, my knowing that the tallest philosopher at University of Johannesburg is wearing a blue shirt would *not* count as self-knowledge, even if I am he, so long as I don't realize that I am he.

Any self-knowledge in the more restricted sense of knowledge about one's mental states (assuming one has that knowledge "from the inside") will be self-knowledge in Perry's sense, which is why Perry's sense is broader. But it's also fair to say that understanding self-knowledge in Perry's sense is required for complete understanding (if such is possible) of self-knowledge in the more limited sense. Suppose a psychotherapist tells me that the tallest philosopher at University of Johannesburg is anxious. This could be knowledge of one of my mental states without properly being self-knowledge, because unless I realize that I am the person so described, I won't really know that *I* am anxious. So, importantly, addressing Perry's concerns is crucial to addressing the more restricted concern of self-knowledge of mental states that has so gripped the philosophical community.

Let's see how Perry classifies the different ways of having knowledge about oneself.

In "Myself and *I*," Perry distinguishes three ways one could have information about—or in some sense "know" about—oneself (Perry 1998c).

1: agent-relative knowledge
2: self-attached knowledge
3: knowledge of the person one happens to be

In the next subsections, I'll explain all three of these. It will turn out that only self-attached knowledge (2) is self-knowledge in the sense we're looking for. But understanding the other senses is required for understanding this one.

Agent-Relative Knowledge (1)

At the beginning of "Thought without Representation," Perry writes:

> I see a cup of coffee in front of me. I reach out, pick it up, and drink from it. I must then have learned how far the cup was *from me*, and in what direction, for it is the position of the cup relative to me, and not its absolute position, that determines how I need to move my arm. But how can this be? I am not in the field of vision: no component of my visual experience is a perception of me. (Perry 1986d: 171, Perry's italics)

This passage describes agent-relative knowledge, which requires "no concept or idea of oneself" (172), even though the information it encodes is partly about the self.

To put it another way: Although visual percepts encode information that is in some sense about oneself—e.g., I have a computer screen in front of *me*—the *self* finds no articulation in those percepts. The seen computer screen is articulated by shape, color, and distance. The wall behind it is similarly articulated. But where is the *self* in the visual percept? Nowhere! Nevertheless, that percept locates the self in his/her environment.

Agent-relative knowledge is the kind of "self-knowledge"[6] that's available to a cat, even though (we assume) cats have no *concept* of the self. When a cat sees a mouse before her, she knows how far to pounce, which means her visual percepts convey how far the mouse is from *her*. We can see this by considering two other proposals for the informational content of the cat's visual perception of the mouse's location. One might say that the cat just sees *where* the mouse is—not in relation to her—simply *where* the mouse is.[7] But unless this "where" is related to the cat, how does she know how far to pounce? The cat has no representation of absolute space, and hence has no representation of herself or the mouse in it. Rather, the cat must see the space around her relation to *her*. A more tempting (but still wrong) proposal is that the cat sees the mouse *only* as spatially related to the environment they are both in—say, as being in a location in the room. Although better, this proposal also fails. For the cat sees the mouse differently depending on where *she* (the cat) is, even if the mouse is in the same spot in the room; the mouse takes up more of the cat's visual field when she is closer to it. It follows that the cat's percepts of the mouse don't just carry information about where the mouse is in the room; they also carry information about how far the mouse is from *her*. Thus, by the same token, they carry information about how far *she* is from the mouse, even though she in no way sees herself. These same points, I believe, would generalize to many other species.[8] In sum, agent-relative knowledge relates one's environment

---

[6] The scare quotes here indicate that this way of having information falls short of self-knowledge in the classic Perryan sense, in which the self is represented or articulated by the self-notion.

[7] This suggestion is of course simplistic. But I did hear it made at a professional philosophy conference.

[8] Yet this humble form of self-knowledge is also critical to all *human* action. Why "all"? The reason is that: "However complex our lives are, everything we do comes down to performing operations on the objects around us—objects in front of us, behind us, above us; objects we are holding; objects we can see" (Perry 1998c: 326–7). Even when I write an email to someone 8,000 kilometers away, I must be aware of how far the keyboard is from me, which requires agent-relative knowledge. Furthermore, many (most?) actions even bypass more sophisticated

to oneself (conversely, oneself to one's environment), without the self's being an articulated constituent of the representations in which the knowledge consists.

But that's not all agent-relative knowledge does. Much agent-relative knowledge consists of information about the relation between the agent and some external object, as I just described. But—crucially—much agent-relative knowledge is *just* about the agent. This tight feeling conveys to me that my shoulder is tense. Just as the notion of *me* is not a constituent of the cat's mouse perception, so the notion of *me* is not a constituent of the shoulder pain. In Perry's terms, there are many "agent-relative roles"[9] occupied by objects in our environment—the keyboard is in the *thing-my-fingers-are-touching* role—and objects in agent-relative roles are sources of agent-relative knowledge.[10] Importantly, one logical relation that defines an agent-relative role is *identity*, in which case the occupant of the role is identical to the receiver of the information.

Thus, an important sub-class of agent-relative knowledge is acquired through what Perry calls "normally self-informative ways of knowing". Pleasures, pains, the feeling of bodily postures, and other sensations are all normally self-informative ways of knowing. They are architecturally guaranteed—guaranteed by the way the organism is set up—to carry information about the agent to the *same* agent, even though the agent him/herself is not an articulated constituent of the representational vehicle of that information. The important feature of this class of agent-relative knowledge that follows from this is that the agent can't be wrong about whom it is about. I can be wrong about whom I see when I see you, but I can't be wrong about whose shoulder hurts when my shoulder hurts. Following Shoemaker, Perry

---

forms of self-knowledge. People often pick up coffee cups and drink without deciding to do so on each act of picking up; rather, having decided to drink already, their bodily movements are adjusted by the varying agent-relative knowledge coming in through visual, proprioceptive, and olfactory faculties.

9 Perry introduces this phrasing in his "Self-Notions" (1990c). That essay also introduces the phrases "normally self-informative ways of knowing" and "normally self-effecting ways of acting".

10 In "Myself and I", Perry distinguishes between basic and derived agent-relative roles. Basic agent-relative roles are defined by relations between agents and things in their nearby environments. My chair is in a basic agent-relative role in relation to me, the thing-I'm-on role. Derived agent-relative roles are occupied by things more distant that relate to us *via* the occupants of basic agent-relative roles. Perry's example is that Bill Clinton is the occupant of the person-I'm-watching role, but this is derived, since the TV must first be in the thing-I'm-watching role. In this paper, since I only talk about basic agent-relative roles, that's what "agent-relative role" refers to in all cases here.

calls this kind of information "immune to error through misidentification".[11]

Why is immunity to error through misidentification so important? It's what enables normally self-informative ways of knowing to ground the self-notion, which is an essential constituent of self-knowledge in the more robust sense, which Perry calls "self-attached knowledge". I explain how this works in the penultimate section. Now let's turn to classifying the other two forms of knowing about the self.

Self-Attached Knowledge (2) and Knowledge of the Person One Happens to be (3)

Why the term "attached"[12] in "self-attached knowledge"? The short answer is that this kind of self-knowledge is attached to the self-notion.

> Ordinarily all one's knowledge about oneself is integrated around a special sort of idea or notion of oneself that we express with "I." While my perception that the beer is in front of me may not require a representation of myself, the information I acquire is immediately integrated into self-attached knowledge, that I might express with "I see beer" or "there is a beer in front of me." And when I read a piece of email that says that John Perry's paper is overdue, I integrate this information into self-attached knowledge, "My paper is overdue," and I realize that it is me that has to get to work. (Perry 1998c: 333)

We can see the importance of the self-notion by looking at two examples that highlight the difference between self-attached knowledge (2), which deploys the self-notion, and knowledge of the person one happens to be (3), which doesn't. The first example is one of Perry's favorites (with help from Ernst Mach), and the second example is mine (with help from Victor Hugo). In each example, the agent starts with knowledge of the person he happens to be, without having self-attached knowledge.

Example 1: Ernst Mach reports that he stepped onto a bus one day and saw a shabby pedagogue at the other end, a man with rumpled clothing and very academic looking (Mach 1914). He thought *that man is a shabby pedagogue*.

---

[11] Sensations of phantom limbs are not a counterexample here. The phantom limb sensation is of something that doesn't exist, but the agent still can't be wrong about whose phantom limb it is. Remember, in "immunity to error through misidentification" it's the *who* (yourself) you can't be wrong about, even though you might be wrong about what it is you sense. Another example (from Perry), on feeling flushed, you might wrongly think you're blushing, but you still can't be wrong about who's feeling flushed.

[12] Note to the reader: The terms "linked" and "attached" have a special meaning for Perry, which I clarify in view of his broader cognitive picture in the section called "Question 3".

Example 2: Jean Valjean is in prison and hears the announcement "24601 will be put in solitary confinement." He thinks *24601 is unfortunate*.[13]

Mach's thought doesn't motivate him to improve his appearance, and Valjean's thought doesn't cause him dread. Yet Mach was just seeing himself in a mirror at the end of the bus, and 24601 is the prison number of Jean Valjean.

Mach and Valjean both had knowledge about themselves without realizing it, which is exactly what Perry calls "knowledge of the person one happens to be". Mach, in some sense, should have been motivated to tidy up, while Valjean should have felt dread. But they didn't, and the problem is a failure of realization. Generally, knowledge of the person one happens to be consists of information about oneself, in which the notion that is *of* the person in question is not the self-notion and is not linked to it either.

To have the needed realization, Mach would have to have a thought that predicates being the man he sees of *himself*, a thought which would involve his self-notion. Similarly, Valjean would have to have a thought that predicates being 24601 of himself. But the sources of images in the mirror are not always clear, even when it's *you*, and numbers used as names are easy to forget, even when it's your number. So not all knowledge about oneself is self-attached.

To attain genuinely self-attached knowledge, Mach and Valjean would have to have thoughts with these structures:

{self-notion} am [man appearing at the opposite end of the bus]
{self-notion} am [24601]

This makes clear why we need an account of the self-notion, for without this we can't make sense of the above thoughts, and those thoughts just are instances of self-knowledge in the required sense. But providing such an account is tricky. One might say that the self-notion is just a notion that refers to the self. But in Valjean's case, [24601] is a notion that refers to Valjean's self, but it's not his self-notion. So the phrase "a notion that refers to the self" won't do as a definition of *self-notion*. Alternately, one might say that the self-notion is the notion one expresses with the word "I". That's true, but not explanatory. In particular, it doesn't tell us why, prior to Valjean's realization, he expresses {self-notion} with "I" but not [24601].[14] And to understand why, we need an account of the self-notion.

---

[13] This is not actually an episode from *Les Miserables*; rather, it's inspired by the combination of my reading Perry and listening to musical theatre.

[14] An interesting question that arises here, although I don't have space for it, is the following. If Valjean both believes *I am not going to solitary* and believes *24601 is going to solitary*,

So explaining the self-notion is still needed. To wrap up this section, this chart summarizes key features of the three forms of self-knowledge.

Table 6.1

	Agent-relative knowledge	**Self-attached knowledge**	**Knowledge of the person one happens to be**
Conceptual structure	Lacks a constituent for the self (i.e. no self-notion), but informationally relates the surrounding environment and the agent him/herself to the agent	Uses the self-notion in predicating a property of oneself (e.g. {self-notion} am *writing a paper on self-notions*)	Uses a notion that, unbeknownst to oneself, refers to oneself (e.g. [24601] *is headed to solitary*)
Form of verbal expression	Will often lack any constituent that refers to the self (e.g., "There's the cup!")	Uses first-person indexicals like "I" (e.g., "I am writing a paper.")	Uses third-personal name or description ("24601 is going to solitary.")
Relation to action	Required for guiding behavior in relation to the immediate environment (e.g. one's cup percepts when one is grabbing the cup)	Used in practical reasoning and decision making, which issues in action (e.g., I desire that {self-notion} *finish writing*; I believe *coffee helps writing*; action consequence: I get more coffee.)	Often mere verbal expression or heuristic behavior (finding out who); cannot be used in ways that treats the subject of the knowledge as the agent (e.g., Valjean won't try to escape *his* solitary confinement on learning of 24601's.).

has he believed a contradiction? The question is of course analogous to Kripke's famous Pierre case, where Pierre believes *Londres est beaux* but also *London is not beautiful*. Perry's solution to that problem, which can be extended to the present case, is found in his "Reference and Reflexivity" (2001B2).

4 Question 2: How Does Self-Knowledge Fit in with the Structure of Knowledge in General?

Understanding the elusive self-notion is critical to understanding self-knowledge. But to understand the self-notion, we must first understand Perry's view of ideas, notions, files, and beliefs in general. In "Knowledge, Possibility, and Consciousness," Perry describes the structure of belief and knowledge with the metaphor of a building.

> Think of the architecture of our beliefs as a three-story building. At the top level are detached files (ideas associated with notions). … At the bottom level are perceptions and perceptual buffers. Buffers are new notions associated with the perceptions and used to temporarily store ideas we gain from the perceptions until we can identify the individual, or form a permanent detached notion for him, or forget about him.
>
> The middle level is full of informational wiring. Sockets dangle down from above, and plugs stick up from below. The ideas in the first floor perceptual buffers and in the third-floor files are constantly compared. When there is a high probability that they are of a single person or thing, recognition (or misrecognition) occurs. The plug from the buffer is plugged into the socket for the notion. Information then flows both ways.
>
> The information flowing up from the perception adds new ideas to the file associated with the notion. … The information flowing down to the bottom level enriches the perceptual buffer and guides my action toward the objects I see and hear in ways that would not be supported just by the ideas picked up from perception. (Perry 2001B2: 120–1)

One central idea of this passage is that the objective knowledge on the top floor needs to be connected to the perceptual information on the bottom floor in order for that objective information to be usable in action. In other words, *I* can't use objective information unless it's attached to something or other on the bottom floor. This gives a clear idea of where to look for the self-notion: the bottom floor. Let's parse a few of the terms Perry introduces.

Constituents of thought, for Perry, are mental entities that can be distinguished by the type of reference they have. Constituents of thought that represent general properties and relations, like the property *being red* or the relation *being a brother of*, are *ideas*; constituents of thought that represent particular things, like *Barry Sanders' garage* are *notions*. A *file* is just Perry's way of talking about the collection of information associated with one's notion of a particular thing. So my file for *Barry Sanders' garage* will have the idea *red* in it if, and only if, I believe that Barry Sanders' garage is red.[15] If I come to believe Barry Sanders has painted his garage, my file

[15] What if I merely imagine Barry Sanders' garage is red, without having a belief on the matter either way? Will I then still have the idea *red* in my file for Sanders' garage? That issue

may then update to have the idea *blue* in it. Finally, a *perceptual buffer* is just a first-story notion/file that tracks information about an object (it could be oneself) through one or more perceptual channels.[16]

What is remarkable about human cognition is that one can have a notion/file of Barry Sanders' garage, without ever having seen it. I may just have read about it in the "Almanac of Detroit Lions Greats" and formed my notion and beliefs solely on that basis. If so, I could find myself in a curious situation: walking through Sanders' neighborhood, I may see the garage, without knowing it's the garage I have beliefs about. On seeing it, I form a perceptual buffer for the garage, which starts to accumulate perceptions and ideas: *pretty*, *red*, *rectangular*, *has room for three cars*, *has a running back in it*, and so on. I then notice a high degree of coincidence between my third-story "detached" file for Barry Sanders' garage, which I formed on reading the "Almanac", and my first-story perceptual buffer.

The moment I say "Aha! That's Barry Sanders' garage!" is the moment, to continue Perry's metaphor, when the plug sticking up from the garage perceptual buffer connects with the socket dangling down from the Barry Sanders' garage file on the third floor. We can notate the structure of my thought as follows (where subscripts indicate: cognitive structure/information source):

$\{\text{This}_{\text{VisualBuffer/Garage}}\}$ is [Barry Sanders' $\text{Garage}_{\text{ThirdStoryFile/Almanac}}$]

Once this judgment is reached, my file [Barry Sanders' $\text{Garage}_{\text{ThirdStoryFile/Almanac}}$] is *attached*; it's linked to—and temporarily contains the contents of—the particular perceptual buffer that I am using to track the garage before me. When I leave Sanders' neighborhood, I'll still have the third-story file, but it won't be attached to any buffer. The file will have been modified to include some things I learned from visual experience. So the file will now have the structure [Barry Sanders' $\text{Garage}_{\text{ThirdStoryFile/Almanac\&Vision}}$]. What's the point of having detached files? First, you wouldn't want to *forget* what you learned. And second, having a detached file, if its information is largely accurate, is useful in shaping anticipations of objects for when one actually encounters them. For example, if I had read in the "Almanac" that Sanders keeps a ferret in his garage, and if I am allergic to ferrets, I'll know to keep my distance, even if the ferret hasn't made its way into my perceptual

---

lies outside this paper. But my own view—not Perry's—is that the idea *red* would still be in the file, just with an imaginative (non-belief) attitude valence.

[16] Places Perry introduces this terminology: Perry 2001B1: 50 and 1998c: 325.

buffer.[17] Detached files can become attached and thereby make their contents useful.

We can already see that the self-notion must be a first-floor notion, but one that in different forms of self-knowledge attaches to files on the third floor. *I* am the entity that perceives and does things in my immediate environment (using first floor information). *Neil Van Leeuwen* is an entity whose name appears on an airplane ticket and is represented in an electronic airline database (third floor information). So knowing the identity that *I am Neil Van Leeuwen* allows the *I* entity that perceives and does things, whenever the airport is my immediate environment, to use the fact that "Neil Van Leeuwen" in on a ticket and in a database.

We are now ready to answer the important question of how the self-notion is constituted.

5 Question 3: How Can We Characterize the Self-Notion?

We've seen that the self-notion is a first-story notion. Let's consider another example that will allow us to characterize it more precisely.

While on the rowing machine in the gym, I notice that someone doesn't smell too good. At this point, I have two linked notions. I have a first-floor olfactory buffer, which tracks the odor. But I also have a third-floor, objective notion/file of the person who doesn't smell too good. Once I leave, I can use this third-floor notion to think about that person, even after the smell is gone, under the description *the person who didn't smell too good*. But now suppose I'm still in the gym and go to a part of it where there are no other people, and suppose I notice the smell persists. Thinking through the possibilities, I come to the thought:

{Self-Notion} am [The Person Who Doesn't Smell Too Good$_{\text{OlfactoryBuffer/PersonWhoSmells}}$]

I learn *I* don't smell too good. I realized this by seeing that *I* was in a part of the gym without other people, while smelling the odor still.

Let's pause on the locution just used: "seeing that *I* was in a part of the gym". I didn't see my*self*. Rather, my visual percepts were of a part of the gym with no one else in it. So why does it feel so natural to use the word "I" in describing what I saw?

The answer is that my visual buffer of nearby surroundings feeds location ideas into my self-notion. This is the crucial point. Part of what it is *to be* a self-notion is to receive location ideas from one's visual or other per-

[17] Let me just stress that this example is purely fictitious. I have no idea whether Barry Sanders has a ferret.

ceptual buffers, including especially the buffers for normally self-informative ways of knowing. More precisely:

> Claim 1: Part of what constitutes the self-notion is that its ideas of the location of the individual it is about *come from* (at least in part) the perceptual buffers of the individual in whose mind the self-notion occurs.

This claim falls under Perry's more general claim about what makes for a self-notion. Recall *Perry's Thesis*: "[S]elf-notions are those that have the special role of being the repository for information gained in normally self-informative ways and the motivator for actions done in normally self-effecting ways" (Perry 2002c: 205). Claim 1 is a corollary of Perry's Thesis, so long as looking around (taking in one's environment through a visual buffer) is counted as a normally self-informative way of learning where one is.[18] We can add:

> Claim 2: Part of what constitutes the self-notion is that its ideas of the bodily state of the individual it is about *come from* (at least in part) the sensations of pleasure and pain in the individual in whose mind the self-notion occurs.
>
> Claim 3: Part of what constitutes the self-notion is that its ideas of the bodily posture of the individual it is about *come from* (at least in part) proprioceptive sensations in the individual in whose mind the self-notion occurs.

We could carry on indefinitely with further corollaries. But the general idea is clear: Part of what makes for a self-notion is its associated file packed with information gained in certain ways; those are not generally ways of information acquisition like reading a book; rather, they include cognitive activities like looking, feeling, and touching—ways of information acquisi-

---

[18] Note the immunity to error through misidentification: you can be wrong about what you see when you look around to learn your location, but you can't be wrong about *whose* location it is you're learning about by looking around. The relevant quotation: "A perceptual state S is a normally self-informative way of knowing that one is φ if the fact that a person is in state S normally carries information that the person in state S is φ and normally does not carry the information that any other person is φ" (Perry 2002c: 204). I do think, to be honest, there is some looseness to what counts as a *normally self-informative way of knowing*. Sometimes Perry seems to characterize this concept in terms of the identity between whom the way of knowing is about and the person who knows in that way. But other times the immunity to error through misidentification is what's important, as is suggested by the quotation in this footnote. The extension of the concept characterized in the latter way is broader than its extension characterized in the former way. But it seems to me that the broader extension is still useful in characterizing the self-notion, which is the main project here, so I focus on that and thus include perceptual states of the external world among the normally self-informative ways of knowing.

tion in which the agent whose situation is learned about is architecturally guaranteed to be the same as the agent learning.

Perry's Thesis includes "normally self-effecting ways of acting". What are those? *Those are ways of acting in which the agent affected by an action normally just is the agent*. Perry gives the example moving one's arm in such a way as to bring a cup to one's lips (Perry 2002c: 205). That's not a way of giving water to someone else; it's a way of giving water to yourself. A host of actions fall in this category: scratching, lifting food to your mouth, holding a flower to your nose, wrapping yourself in a blanket, and so on. These are all quite basic actions, and to perform them, I do *not* conceive myself under a detached, third-personal notion, like [the guy in office 735A with a blue shirt]. Rather, I find that *I* {self-notion} desire water.

Let's flesh this picture out. Desires are mental structures that motivate action. As such, they are typically structures that represent something as happening—often they represent a certain agent as getting something. So, often, a desire is structured as: {agent notion}/[agent notion] gets {object notion}/[object notion].[19] Now we ask again: How can we define *self-notion*? We've already identified a class of actions independently: The normally self-effecting ways of acting. So, along with Perry, we can propose: The self-notion is what occupies the agent notion role in those desires that motivate normally self-effecting ways of action.

This view could yield hundreds of additional corollaries to Perry's Thesis, but I'll just mention three.

> Claim 4: Part of what constitutes the self-notion is that the self-notion occupies the agent notion role in desires that issue in lifting drinking vessels to one's lips.
> Claim 5: Part of what constitutes the self-notion is that the self-notion occupies the agent notion role in desires that issue in scratching oneself.
> Claim 6: Part of what constitutes the self-notion is that the self-notion occupies the agent notion role in desires that issue in wrapping oneself in a blanket.

These Claims, of course, only apply to agents prone to doing the actions in question and for whom those actions would be normally self-effecting; basically, humans. We could, by way of contrast, imagine aliens who never drink or wrap themselves in blankets, and to have their itches scratched they call the central Martian scratching agency. For such aliens, other claims would be needed to define the self-notion, ones that appeal to the normally

[19] The relevant points will go through for other desire structures.

self-effecting actions *for them*. But Claims 4–6 do well for human self-notions.

The picture Claims (1–6) paint is of the self-notion as the informational node linking normally self-informative ways of knowing (and agent-relative knowledge more generally) to normally self-effecting ways of acting. Because these ways of knowing and acting can be defined independently of the self-notion, by way of agent-relative information and the identity relation, this way of defining *self-notion* is clearly not circular.

Let's use this picture to address our example of Jean Valjean. Valjean has more than one notion that refers to himself. Only one of them is the self-notion. Which is it? Why isn't the [24601] notion the self-notion? We now have an answer. Lacking the appropriate cognitive links, [24601] is not the repository of normally self-informative ways of knowing and is not implicated in the motivation of normally self-effecting ways of acting. Before Vajean realizes he's 24601, seeing a fence in font of him doesn't feed into Valjean's 24601 notion. *His* fence percepts don't give him ideas about where 24601 is. Nor does Valjean's (seemingly altruistic) desire that 24601 be well nourished cause Valjean to lift food to his *own* mouth. But now suppose Valjean has another notion that refers to himself; call this notion V. Suppose further that when Valjean sees a fence, the idea of facing a fence is fed into Valjean's V notion. He thinks: {V} is/am [facing a fence]. Furthermore, when Valjean feels a toothache, the idea of having a toothache is put in the V file. And whenever Valjean desires that V have a drink and there is a cup of water nearby, this causes Valjean's arm to lift the cup to his own lips. If V has all these relational properties—in short, if V satisfies Perry's Thesis and the corollary Claims 1–6—then V *just is* Valjean's self-notion.

So far, the self-notion seems primitive, a repository for basic information and motivator for basic actions. But the self-notion *can* link to third-story objective knowledge. The plug sticking up from the self-notion can connect with sockets dangling down from the third floor, sockets that dangle down from notion/files such as [24601] or [the shabby pedagogue on the bus]. In Valjean's case, {V} becomes linked to [24601]. When that happens, if Valjean knew 24601 was going to solitary, he would realize *he* is going to solitary. He comes to anticipate as host of normally self-informative, agent-relative information: seeing darkness, feeling hunger, feeling a damp stone floor. And he may try to hide a bottle of water under his shirt, anticipating he'll need it to perform the self-effecting action of bringing a drinking vessel to his lips. In short, his previously detached knowledge about 24601 becomes suddenly pertinent to how he sees, feels, and acts on the environment immediately around him. The self-notion is not

merely the informational node linking agent-relative knowledge to normally self-effecting action; it is also the node that can link such ground-level epistemic and pragmatic ways of knowing and acting to objective knowledge that happens to pertain to the self. As Perry puts it, "My self-notion can be both tied to an epistemic/pragmatic relation and serve as a permanent file for myself" (Perry 2002c: 206).[20]

6 Conclusion: What Perry Accomplishes

What does Perry's theory of self-notions, self-belief, and self-knowledge accomplish? I think the biggest accomplishment is that the theory comes as close as possible to characterizing what it takes for a cognitive system to have self-knowledge, without resorting to descriptions from the inside—or first-personal, phenomenological descriptions. It's true that Perry mentions sensations, like feeling flush, in exemplifying normally self-informative ways of knowing. But the phenomenology of those states is not what's important to Perry's project. Rather, what's important is (1) the logical relation of *identity* between the person those states carry information about and the person who is in those states and (2) the architecture of the cognitive system that ensures that identity.

To summarize, the first theoretical step[21] is independently to classify ways of acquiring information and forms of action. This yields the categories of agent-relative information, normally self-informative ways of knowing, and normally self-effecting ways of acting. Once these forms of information reception and action have been classified, Perry can use them noncircularly to define the self-notion; the self-notion is the repository of information acquired from the normally self-informative ways of knowing and is a constituent in desires that motivate self-effecting ways of acting. Filling these roles, the self-notion is also able to relate the self of the here-and-now, which must drink water and avoid tree stumps, to normally de-

[20] Note that Perry's way of looking at self-knowledge makes salient a logically possible form of amnesia that, as far as I know, never occurs. One could remember all of one's biographical information, but still have forgotten which person, of the persons one has knowledge about, one is. For example, if I had this form of amnesia, I might still remember Neil Van Leeuwen's history, but still not know whether I am he or I am Albert Newen, some of whose history I also know, albeit to a much lesser extent. I think it argues in favor of Perry's view that this form of amnesia never happens. For one thing that must prevent it from happening is that some of my knowledge of Neil Van Leeuwen's history comes in the form of perceptual memories, like visual memories. Saying that such memories are first-personal and saying they encode agent-relative information is almost saying the same thing, which argues in favor of seeing the agent-relative information as self-notion constituting.

[21] In order of logical priority, not necessarily in order of presentation.

tached information, like information about which stream has fresh water and which trees attract poison oak.

The next theoretical moves are clear. The concept of a self-notion can be used to characterize self-belief—beliefs in which the self-notion is a constituent—which can be used to characterize self-knowledge.

Thus, Perry manages to give a theory of self-knowledge that starts with very humble origins. The categories of agent-relative knowledge, normally self-effecting actions, and normally self-informative ways of knowing apply just as well to cats and dogs. Dogs scratch their ears, just as humans do, and cats feel their hair standing on end, just as humans feel their hearts race. But Perry shows that from these humble origins is constructed the core, the self-notion, of what is often regarded as humankind's singular cognitive achievement: self-knowledge.

Acknowledgments

I'd like to thank Wes Holliday, Thad Metz, Albert Newen, and Raphael van Riel for comments on earlier drafts of this paper. Most of all, I'd like to thank John Perry, without whom I *couldn't* have written this!