THE SELF THE SELF, SELF-KNOWLEDGE AND SELF-NOTIONS

By John Perry

[This essay is Chapter 10 of *Identity, Personal Identity and the Self*. Indianapolis: Hackett Publishing, 2002]

1. "SELF" AND SELF

The English expression "self" is a modest one; in its normal use, it is not even quite a word, but something that makes an ordinary object pronoun into a reflexive one: "her" into "herself," "him" into "himself" and "it" into "itself." The reflexive pronoun is used when the object of an action or attitude is the same as the subject of that action or attitude. If I say Mark Twain shot *himself* in the foot, I describe Mark Twain not only as the shooter but as the person shot; if I say Mark Twain admired *himself*, I describe him not only as the admirer but as the admired. In this sense, "the self" is just the person doing the action or holding the attitude that is somehow in question. "Self" is also used as a prefix for names of activities and attitudes, identifying the special case where the object is the same as the agent: self-love, self-hatred, self-abuse, self-promotion, self-knowledge. When we say "the same" and "identity" in these contexts, we mean that there is *only one thing*. The way I use "*A* and *B* are identical, " it means there is just one thing that both *is A* and *is B*. Mark Twain and Samuel Clemens are identical because there is just one fellow that was Mark Twain and was Samuel Clemens. When I use the word "identity" in some other way, I'll put it in scare-quotes.

Given the meaning of "self," one might expect the phrase "the self" to simply be an alternative way of saying "the person," and sometimes it is used this way. In fact the term is often appropriated for various inner agents or principles that are thought on various philosophical and religious views to be necessary for

consciousness, knowledge, freedom, or personal identity. The phrase is also often used to refer to the most central parts of the concept a person has of himself or herself.

In different disciplines, somewhat different things are regarded as most central. In psychology "the self" is often used for that set of attributes that a person attaches to himself or herself most firmly, the attributes that the person finds it difficult and/or disturbing to imagine himself or herself without. The term "identity" is also used in this sense. Typically, one's gender is a part of one's self or one's "identity"; one's profession or nationality may or may not be.

In philosophy, the self is the person considered as agent, knower, subject of desires and conscious subject of experience. These are philosophically the most central parts of a person's self-concept: I am the person doing this, knowing this, wanting this, having these sensations and thoughts. It is this concept ourselves that that is extended through memory and anticipation and forms the basis of personal identity. I am the person who *did* this, and *will* to that; I am the person who *had* this experience, and wants to have it again. If the present thought of future reward or punishments is to encourage or deter me from some course of action, I must be thinking of the person rewarded as me, as myself, as the same person who is now going to experience the hardships of righteousness or not experience the pleasures of sin to gain this reward.

Given the meaning of the particle "self," and the nature of our self-concept, a reasonable hypothesis is that *the* self, in the philosophical sense, is simply *the* person who is the knower, agent, subject of desires, and possessor of thoughts and sensations. It seems this same self, the knower, agent, and conscious subject comes up in many quite mundane transactions, and turns out to be the person. If I pick up the cake and shove it in this mouth rather than that one, isn't it because I think it is *my* mouth, and so it will be me, the very same person who picks up the cake, that will have the pleasure of tasting it? Isn't identity of the metaphysical agent, the ultimate locus of reward and punishment, simply being

the same person, simply identity? Isn't what I worry about, when I worry about going to prison or going to Hell simply that the person to be punished and I are one and the same ---identical, without scare quotes? If so, this self, the identity of which is at the bottom of every action, and involved in every bit of knowledge, the self of the philosophers, is simply the person who does the action, and has the knowledge---not anything more, or less, mysterious.

A straightforward view of the self then is be that the self is just the person, and that a person is a physical system, with the unity physical systems can possess, not a unity based on some other inner agent and perceiver, or mysterious principle. This view has been challenged on (at least) two fronts. First, the nature of freedom and consciousness has convinced many philosophers that there is a fundamentally non-physical aspect of persons. I'm not going to talk about these issues. As to freedom, the problem is large and complex, the issues are familiar to most philosophers, and I don't have anything new to say about it ---- although I do have hopes of some day having a good idea. The arguments in favor of immateriality of the mind or self do not have as strong a hold on the philosophical community as they once did. While there are many philosophers who think that mental properties cannot be fully reduced to physical or material properties, most such philosophers would allow that these are properties of a physical system, rather than an immaterial self. I've considered the issue of whether the nature of subjective experience shows that not all of our properties are physical at length in *Knowledge, Possibility and Consciousness* (2001).

The second challenge stems from puzzling aspects of self-knowledge. The knowledge we have of ourselves seems very unlike the knowledge we have of other objects in several ways, and this has led some philosophers to rather startling conclusions about the self. In his *Tractatus*, Ludwig Wittgenstein tells us that "I am my world," and that "the world is my world" (Wittgenstein, 1961, 5.63,5.641). This should lead us to the rather surprising conclusion, that I am the

world, or that at least Wittgenstein was. He draws at least one conclusion that would follow from this; he says "...at death the world does not alter, but comes to an end."

The contemporary philosopher Tom Nagel has been led to a possibly less radical but still quite dramatic view. According to Nagel, when he says "I am Tom Nagel," at least in certain philosophical moods, the "I" refers to the "objective self," which is not identical with, but merely contingently related to, the person Tom Nagel. This self could just as well view the world from the perspective of someone else other than him (Nagel, 1983). I discuss Nagel's view at length in Essay 11. Here we will examine puzzling features of self-knowledge that give rise to such views.

2. *SELF-KNOWLEDGE*

"Self-knowledge" seems to have a straightforward meaning: cases of knowledge in which the knower and the known are identical. But this doesn't seem sufficient. In a footnote to his book *The Analysis of Sensation* (Mach, 1914, 4n.), the philosopher Mach tells of getting on the end of a bus, and seeing a scruffy unkempt bookish looking sort of person at the other end. He thought to himself,

(1) That man is a shabby pedagogue.

In fact, Mach was seeing himself in a large mirror at the far end of the bus, of the sort conductors used to use to help keep track of things. He eventually realized this, and thought to himself:

(2) I am that man

(3) I am a shabby pedagogue.

Now consider Mach at the earlier time. Did Mach have self-knowledge? In our straightforward sense, it seems that he did. After all, he knew that a certain person was a shabby pedagogue. Furthermore that person was, in fact, him. The knower and the known were the same. But this case isn't really what we have in mind when we talk about self-knowledge. Self-knowledge is

something Mach really only had when he got to step (3), when he would have used the word "I" to express what he knew.

Self-knowledge in this restricted sense seems peculiar. First, it seems "essentially indexical." Statement (3) expresses self-knowledge because of the word "I"; it is hard to see how Mach could have expressed self-knowledge without using the first-person. If he said "Mach is a shabby pedagogue," he would be only claiming to know what everyone else may have known---something he could have learned by reading the papers, even if he had amnesia and didn't know who he was, or that he was a shabby pedagogue. It doesn't seem that there is any objective characterization *D* of Mach, such that knowing that *he* is a shabby pedagogue, amounts to knowing that *D* is a shabby pedagogue (Castañeda, 1966,1968; Perry, 1990, 1993).

Secondly, we seem immune to certain sorts of misidentification with respect to self-knowledge. If we learn, in certain ways, that someone is in pain, then we cannot miss the fact that it is we who are in pain. That is, if Mach discovers that he has a headache in the ordinary way that a person discovers they have a headache, he can scarcely be wrong about *who* has the headache, if the range of choices is "I/you/that man, etc." Of course he can be wrong if the range of choices is "Mach/Freud/Wittgenstein," etc., for he might not realize which of those people he is, if he has amnesia. (Shoemaker, 1984)

Thirdly, self-knowledge seems to play a unique cognitive role. If Mach desires that *he* do so and so, and believes that *he* can do so and so by executing such and such a movement, then he will execute that movement without further ado (Perry, 1990).

While (3) expresses self-knowledge, (1) does not. And yet (1) is, in a perfectly clear sense, a case of Mach believing something about himself. Mach implies that he was, in fact, a shabby pedagogue. It is because he was a shabby pedagogue that we take the belief expressed by (1) to be true. If it is Mach's being a shabby pedagogue that makes (1) true, then (1) was about Mach, and

expressed a belief about him. Nevertheless (1), unlike (3), is not an expression of self-knowledge.

I shall sometimes use the term "self-belief" rather than "self-knowledge." Although "self-knowledge" is more familiar, it is somewhat misleading since the distinction between knowledge and mere belief is orthogonal to the issues I discuss. I take beliefs to be complex cognitive particulars, that come into existence as a result of perceptions, inferences and other events, and influence the occurrence and nature other beliefs and actions. I assume that two beliefs are involved here, one that Mach acquired when he stepped on the bus, and one that he acquired a bit later, when he figured out that he was looking at himself. I want to understand the difference between those beliefs. It is not sufficient, for this purpose, to note that (3) contains the word "I" where (1) contains the words "that man". This is why (3) is an expression of self-belief and (1) is not. But I want to know why the belief thus expressed is a self-belief.

3. *BELIEFS*

My account will presuppose a fairly common sense view of beliefs. The mind has ideas of things, properties and relations. I'll call ideas of things "notions." A belief is a complex in which an idea of a property or relation is associated with the appropriate number of notions of things. The content of a belief is that the things the notions represent stand in the relation that the idea represents. A number of beliefs with a common notion is a *file*.

The function of beliefs is to retain information picked up through perception, to formulate hypotheses, to allow for the combination and comparison of beliefs and the formation of new beliefs through inference and to motivate actions that will promote one's desires if the beliefs are correct. The object an idea represents depends on its role in psychological system within which beliefs play their role and the way that system is set into the wider world. When Mach stepped on the bus, he formed a notion of the man he saw stepping on the bus. The fact that this was a notion---an idea of a thing--- and not some

other kind of idea depends on the way it functions internally; the fact that it was of a certain man (Mach, as it turned out) depends on external circumstances. If Mach had been looking at some other man, the very same idea would have been a notion of that other man.

We have unlinked notions of the same thing when the external factors that determine which thing the notions represent happen to make them represent the same thing, although there is nothing in the notions themselves, or the ideas associated with them, that reflects this identity. The beliefs involving those different notions of the same thing can function independently. They can arise at different times, and can affect actions in quite different ways. A student who hears of Tully from a Classics Professor and Cicero from a Philosophy professor may believe that Cicero was a philosopher, but not be sure whether Tully was. Cases of this type have been examined extensively in the philosophy of language, usually known "morning star/evening star" cases in honor of the alleged ignorance of the Babylonians that the same heavenly object (namely, Venus) was both (Barwise and Perry, 1999; Crimmins and Perry, 1989; Crimmins, 1992).

It doesn't take long dead philosophers or distant planets to generate such cases. I could easily have two unlinked notions of you, one formed as a result of reading articles by you, one formed as a result of seeing you in the library of my university, where you happen to be visiting without my knowing it. When I read your latest article, the first notion will become associated with new ideas based on what I read. When I see you in the Student Union, the second notion will become associated with new ideas based on what I observe. I have two clusters of beliefs, or files, about you, each consisting of all the ideas associated with one of the two notions. It is theoretically possible for there to be two unlinked notions of the same object, which are associated with the very same ideas, so that a person has two exactly corresponding files for one object. But this seems unlikely. To have two unlinked notions of the same thing, we must

have interacted with it in different ways or circumstances, or at least at different times, and some of the associated ideas are likely to reflect these differences.

Beliefs and files of the same object may motivate quite different actions. When I interact with you, my behavior will be guided by one file or the other, depending on the situation. I know how to write to you, as the author of the articles I have read, for they include your name and a department where you can be reached. Perhaps I write you a letter praising you for your sensitivity and clarity of thought. I know how to speak to you, as the person I have seen in the library, for the file corresponding to my first notion contains information about where you hang out and what you look like. Perhaps, based on my observations of you around campus, where I have seen you go out of your way to be kind, I also think you are sensitive. This leads me to associate the campus-based notion with the idea of sensitivity. This doesn't lead me to write you a letter. There is no information in the campus-based file about your name or address. I just walk up to you one day and compliment you on your kindness and sensitivity. As long as the files remain unlinked, the information in one will not affect the actions guided by the other. For example, I won't call you by your name when I see you, I won't mention seeing you on campus if I write you.

In this example, I have two beliefs with the same content: that you are kind. The beliefs have quite different causal roles. This is explained by the different notions involved in the beliefs, and the other different ideas with which those notions are associated. While the two beliefs have the same content, the files of which they are a part do not. What unifies a file, and makes two beliefs about the same person relevant to one another, is not that they are about the same person, but that they contain the same notion or linked notions. One can have a file made up of a lot of information about different individuals mistakenly associated with a single notion, just as one can have two notions where there is only one object.

The phenomenon of having two unlinked notions of one individual is very common and doesn't require unusual circumstances or unusually confused people. I see my friend Al limping towards me, but cannot yet recognize him; I form a notion of this person. At that moment I have two unlinked notions of Al. Certain of my beliefs about Al I have twice over, such as that he was a man. Others I have in one file but not in the other, such as that he has a limp. I accumulate information about him as he gets nearer; finally I recognize him as Al. At that point the notions become linked; the newly acquired perceptual information combines with the old information and I say, "Why are you limping, Al?" If the identification is tentative, the notions may retain their identity; if not, they may merge and become one.

4. SELF-IDEAS AND SELF-NOTIONS

Mach's confusion is a special case of this sort, where the person he comes to recognize is himself. At the beginning of the episode, Mach had two notions of Mach. One he acquired when he stepped on the bus and saw what he took to be a man at the other end. The other is a *self-notion*, the sort of notion that is usually involved in his beliefs about himself. Beliefs involving one's self-notion have a special role in one's cognitive life, and we usually reserve the term "self-knowledge" for knowledge involving beliefs of this sort. But we need to say more about this special role.

The natural place to look is the ideas with which self-notions are associated. Consider the self-idea, -- the idea we would express as "being me". The notion involved in Mach's first belief isn't associated with this idea, while the one involved in his latter belief is. But which idea is this? We cannot identify the idea by the property it represents. Mach has two ideas of the property of being Mach, one which he would express with "being me" and one which he would express, directing his attention towards the man he sees, with "being him". The former is the self-idea, but why? What makes one of the ideas that represents the property of being identical with Mach his self-idea, and not the other? We might

suppose that the self-idea is a complex idea, composed of the idea of identity in association with the self-notion. This seems plausible, but now we have just gone around in a circle: what is special about the self-notion is that it is associated with the self-idea; what is special about the self-idea is that it has the self-notion as a constituent.

Another approach is to characterize self-notions semantically. We can think of a self-notion on the analogy of the indexical "I". Just as utterances of "I" stand for their utterers, these special self-notions would be of the thinkers to whom they belong. Such notions, and any other notions that were linked to them, would be self-notions. Beliefs with self-notions as constituents would be self-beliefs. The notion Mach acquires when he looks at the fellow in the mirror is not a self-notion at the beginning of the episode, but becomes one when he recognizes himself, and links his "that man" notion to his self-notion. This characterization leaves a question unanswered, however. After stepping on the bus, Mach had two unlinked notions of himself. Which of these should be characterized as a self-notion? They both represent Mach, so that can't be the difference between them.

An analogous question about language would be why we take "Ernst Mach" to be a name of Mach, and "ich" to be the first person in Mach's idiolect. They both stand for Mach, so that isn't the difference between them. The answer to this question lies in the different ways Mach uses these terms in communication, both as speaker and as listener. He takes every utterance containing "Ernst Mach" to be about him, while he takes utterances containing "ich" to be about their speaker.[i] This is suggestive, but doesn't quite get at the difference between Mach's two notions of himself. These notions are not devices for communication. They are not public; Mach does not produce them as an aid to securing recognition by others of his communicative intentions. Rather, they are parts of a system for the pick-up, retention, analysis, discovery and utilization of information by an individual. To explain what we mean by self-

notions, and how Mach's two beliefs differ, we need to explain the role of these notions and beliefs in this informational system.

5. *EPISTEMIC/PRAGMATIC RELATIONS AND R-NOTIONS*

Just as there is a special way of thinking about the person you are, there is a special way of thinking about the place you happen to be in, the way you think of the place you call "here".[ii] Without realizing it, I could be in, say, Grand Island, Nebraska, at the same time I was watching a video of Grand Island scenes. When I thought, "that city looks like a fun place to visit," as I watched the screen, I would be thinking about Grand Island, but not as the place I was at, not as "here." Thinking for a bit about what is involved in thinking about a place as "here," will be helpful in seeing what is involved in thinking about a person as "I".

Suppose you are traveling, and wake up in a hotel room in Grand Island. You look out the window and see rain. So you grab your umbrella before departing from the hotel. Here is a simple case of using information acquired perceptually to guide action. When you look out the window, you get information about the weather in Grand Island. And when you depart from the hotel, your decision to take an umbrella is vindicated because of the weather in Grand Island. The fact that the place whose weather you learn about when you looked out the window is the place whose weather determines whether you need an umbrella is crucial to the success of your use of the information. How do you have to think about Grand Island, to facilitate this use of information to guide action?

One possibility is that you think about Grand Island via its relation to you--as the occupant of an agent-relative role. You think "It is raining here" or "It is raining in this city". A second is that you think about Grand Island via some attribute that is independent of its relation to you, such as its name. The appropriate expression of your thought is "It is raining in Grand Island." These different ways of thinking correspond to different beliefs you might hold,

independently of one another. You look out the window and see rain: It's raining here. You watch TV and hear the reporter on the Lincoln station say "It is cloudy in Omaha, sunny in Lincoln, and raining in Grand Island": It's raining in Grand island. You could acquire either belief without the other, if you had forgotten that you were in Grand Island.

Let's imagine that you have a very poor memory, and keep track of information by using 3x5 cards. You have a number of these for the various cities you frequent, including one for Grand Island. When you hear the news report you take out this card and jot "rain" on it. But we are supposing you have forgotten where you are. When you look outside and see rain, you don't know which of the city-cards to write "rain" on. So you take out your "here" card and write "rain" on it. Call this card the "here-buffer". Information accumulates on this card: Grand Island Chevrolet and Izuzu is on the corner (here). The Grand Island Hotel is across the street (from here). The Grand Island Rotary meets (here) for lunch every Wednesday. At last you figure out that you are in Grand Island. When you do, you transfer the information from the here-buffer to the Grand Island card, which already has "rain" on it. At that point the cards would not only in fact contain information about the same place, but be recognized by you as doing so. Perhaps you put a rubber band around them to help you remember which city you are in. The cards are linked. They not only refer to the same place, this co-reference is reflected in the way you have them organized.

As time passes, you have to update your cards in various ways. But the relatively permanent features you have noted on your Grand Island card, such as "on Interstate 80", "has an interesting museum" do not need to be changed just because you move on to Kearney or North Platte. This card is of Grand Island, whether you happen to be there or not. Your "here-buffer," on the other hand, should be erased, and unlinked from your Grand Island card.

Swithching back to our little model of the mind, let's suppose there is a notion permanently associated a self-notion and the idea of being-at: the place

I'm at. Thois is the here-buffer. For the person who doesn't know where they are, the only here-notion will be a here-buffer. If they have this buffer linked to a permanent notion for a place, that notion will be a here-notion as long as the link is in place.

Note that when the information that it was raining was only in your here-buffer, not linked to your Grand Island notion, you decided to take an umbrella. If it's raining *here* I need an umbrella, no matter where *here* is. What one needs to know, to determine whether rain in a given city provides a reason for taking an umbrella is whether that city is the city one is in. You know that the city you see out your window is the one you are in, without knowing which city it is. On the other hand, having the information that it is raining in Grand Island from the radio, so it is associated with your Grand Island notion but not your here-buffer, will not suffice to motivate taking an umbrella.

There are ways of getting information about the city one is in quite independently of which city it is. And there actions the success of which depends on the conditions prevailing in the city one is in, quite independently of which city it is. One can be motivated by information picked up in these ways to perform these sorts of actions, without knowing where one is. Looking outside one's window is a way of finding out what the weather is like in the city one is in, whichever city that happens to be. And taking an umbrella is an action that will be a good idea, if it is raining in the city one is in, whatever city it happens to be. I shall say that there are "normally here-informative" ways of getting information, and "normally here-dependent, here-directed and here-effecting" ways of acting. It will be reasonable for normally-here informative ways of getting information about cities to motivate normally here-dependent actions whose success depend on that information. That's a theory-laden way of saying that reasonable people take umbrellas when they see rain out the window, even if they don't know which city they are in.

(Taking an umbrella is a "here-dependent" action because its success depends on how things are here. It's successful if it is raining, if not, an unnecesary burden. It is "here-directed" because it is your situation here that your are going to change by doing it. And it is here-effecting because it is your situation here that it will effect. Some times one of these words is more appropriate than others, so I won't always repeat the whole phrase in what follows.)

I shall call relations between an agent and another object -- including places, material objects and other persons -- that support such special ways of knowing and acting, "epistemic/pragmatic relations." The relation of *being at*, that holds between people and places, is an epistemic/pragmatic relation. There are many others. There are special ways to know about the material objects and people *in front of* one (open your eyes and look, reach out and touch), and special ways of dealing with them. There are special ways to know what a person is saying when *are on the phone* (listen to the sounds coming out of the ear piece) and special ways of saying things to them (speak into the mouthpiece). Where *R* is an epistemic/pragmatic relation, we may speak of "normally *R*-informative ways of perceiving'' and "normally *R*-directed/dependent/effecting ways of acting''.

We are all masters of hundreds of such ways of gaining information about things, and dealing with things. They allow us to gain information about things and deal with things without having any way of identifying them independently of their relation to us. They allow us to interact with individuals we know about, once we determine or bring it about that they stand in an epistemic/pragmatic relation to us. When you call, I use such methods to accumulate information about who is talking to me, until I figure out that it is you. Once I realize it is you I link my "on the phone" buffer and my permanent notion for you, and combine the information in them. I know you want to know what Elwood said last night, and I know I can tell you by talking into the mouthpiece, so I do.

The informational role of an *R*-notion is to serve as the normal repository for information gained in normally *R*-informative ways, and as the normal motivator for normally *R*-effecting and *R*-dependent actions. The information I pick up by looking around me will, normally, become associated with my here-notions. The beliefs involving these notions will motivate actions like taking an umbrella, whose success depends on the weather around me. The information that motivates a normally here-dependent action need not have been obtained in a normally here-informative way. If you know you are in Grand

Island, you may take an umbrella because you heard on the radio that it was raining there. The action motivated by information gained in a normally here-dependent way may motivate actions whose success is not normally here-dependent. Seeing that it is raining, you may tell someone on the phone that it is raining in Grand Island, a statement that would be as true if you said it standing in Valentine or Ainsworth.

"Not knowing who someone is" usually amounts to having an *R*-buffer and a permanent file that are unlinked. There are two sorts of cases. In the example above about my friend Al, I had the buffer and need to pick the right file. But suppose instead that Al is a philosopher I have read and written to; he is in a room full of philosophers and I want to talk to him, but don't know what he looks like. I've got the file, but need to pick the right buffer. A calendar entry, with a date and some appointments by it, is like a permanent file. It doesn't help if you don't know what day it is--if you don't know whether to think of the day as "today" or "tomorrow" or what. But it would be equally frustrating to be in the position the cartoon character Ziggy was once depicted as being in: he rips off one page on his calendar and reads "the next day" on the next page. We want calendars to give up objective representations of days so that we can use them to organize information objectively.

We might think of our notions as forming a multi-leveled system. At the top level are notions that are completely, or at least maximally, independent of

relationships to us. These are "objective" representations. The lower levels contain buffers for various relationships to us, associated with various epistemic/pragmatic relations, of increasing specificity. These are the buffers. We pass information up the levels, as we gather information about objects in epistemic/pragmatic relations to us, recognize them and store the information in ways that are more independent of our relationships. We pass information down the levels when we recognize an object and act on it ways that depend not just on its present relation to us, but other properties about which we have gathered information in the past.

When we think of beliefs, we are usually thinking about information stored at the relatively high levels. In fact, it is difficult to describe links between levels if we confine ourselves to "believes" and its cognates, as any philosopher who has dealt with the puzzles from the philosophy of language is aware. We have an additional vocabulary, including "recognizes," "takes to be," and "identifies" to describe linking. For objects and persons with which we are familiar, we have relatively rich permanent files, and it is the contents of these files that primarily count as our beliefs about the thing or person in question. Such beliefs provide the extra or incremental information we have to bring to bear on our interactions with these objects and persons, in addition to what we perceive about them at the time of a given interaction.

6. SELF-NOTIONS AS R-NOTIONS

I believe what is special about self-notions is that they are the normal repository of normally self-informative ways of perceiving, and the normal motivator of normally self-dependent ways of acting. Identity is an epistemic/pragmatic relation.

We might call the example about Mach a "Castañeda example," after Hector-Neri Castañeda, who introduced a number of examples of this sort, and insightfully analyzed them (Castañeda, 1966, 1967, 1968; see also Perry, 1983). They typically involve perceptual states that are not normally self-informative in

the sense I am using the term, but nevertheless in fact carry information about the person that is in the state. A person writes a biography of a war-hero who was missing in action after a certain battle; in fact the person is the war-hero, but doesn't know it, because of amnesia sustained in the battle. A philosopher asks friends about the new editor of *Soul*, not realizing he has been named the new editor. The state Mach was in, when he saw the man in the mirror, was the sort of state one is usually in when one sees that someone else, standing at some distance from one, is shabbily dressed, not when one sees oneself to be shabbily dressed. (Or perhaps he realized that the person he was seeing was reflected in a mirror. Seeing someone in a mirror a short distance in front of one is a normally self-informative way of seeing, but seeing someone in a mirror at the end of a street car is not.)

Contrast with such cases what we might call "Shoemaker cases." Sydney Shoemaker has emphasized that we often find things out about ourselves in ways that are "immune to mis-identification" (Shoemaker, 1963, 1970a; see also Evans, 1982, especially sections 6.6 and 7.2). Suppose you are at a party. You bend over to pick something up, and hear the ripping sound that is characteristic of trousers splitting. Then you feel a hot flush in your face. You are aware that you are blushing. Now who is it, of whose blushing, you are aware? We are almost inclined to say that the question makes no sense. It is of course your own blushing of which you are aware. It's not that you cannot be aware of the blushing of others. You can see them blush. But you can't feel them blush; you can't come to know someone else is blushing, in the way that you typically come to know that you are.

Shoemaker emphasizes that immunity to mis-identification should not be confused with incorrigibility or even privileged access, although they often go together. Compare blushing with being embarrassed. It seems that there is a way of knowing that one is embarrassed, the normal way, which is immune from mis-identification, privileged, and at least close to incorrigible. In the pants-

splitting episode I can't be wrong that I am the one that is embarrassed (immunity); I know this in a way that is more direct and error-free than anyone else can (privilege), and perhaps I can't be wrong about it (incorrigibility). But I can be wrong about whether I am blushing. I may know that I am embarrassed, but be mistaken in thinking that I am blushing. I may not be in as good a position to tell if I am blushing as someone else, who can see my face redden. So my judgement that I am blushing is neither incorrigible nor privileged. But this judgement cannot be wrong because the person I take to be blushing, is blushing, but is not me. Feeling one's face flush is a corrigible way of finding out that one is blushing; but it cannot be used to find out whether someone else is blushing, so if there is a mistake it won't be about *who* is blushing, but only *whether* they are.

It is the way of finding out, not what is found out, that is immune to mis-identification. One could look in a mirror, and think that one saw oneself blushing, although it was someone else. I could believe that I am blushing, and believe it on the basis of observing someone blush, but be wrong about who it is.

These ways of knowing that are immune from mis-identification are, I suggest, just a special case of "normally *R*-informative ways of knowing." 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 the information that that the person in state *S* is φ and normally does not carry the information that any other person is φ. Identity is an epistemic/pragmatic relation. Feeling one's face flush is a way of registering the information that the person identical with the feeler is blushing. Feeling hunger is normally a way of detecting that one's own stomach could use some filling. The feeling of needing to urinate is normally a way of knowing that one's own bladder is full. In each case, someone else can determine the same thing, using a different technique. Perhaps you can see me blush even when I am not aware that I am. Perhaps you know that my stomach is full, having noticed what I have put into it, while I am still in that charming interval between being full and feeling full. Parents often are better judges of how full

children's bladders are than the children themselves are. But you cannot (normally) know that I am blushing, or full, or need to urinate in the way that I do.

Why do I say normally? There are some cases where this qualification is clearly required. Think about watching your hands as you type or play the piano. There is a characteristic way of seeing one's own hands and limbs and torso, a way in which one hardly ever sees anyone else's hands or limbs or torso. Yet when we learn to play the piano, or type, our teacher's hands could conceivably be mistaken for our own; one sees them in the same way as one sees one's own hands. Of course one would quickly spot the mistake, since however similar one's teacher's hands are to one's own, one cannot move them like one moves one's own.

Suppose that a way is developed to repair spinal column injuries by using an external shunt, which connects the column below the injury to the brain stem. One can imagine the shunt having an external connection for some reason. One could go on to imagine that there was enough similarity among people that one person's shunt could be plugged into another person's brain stem receptor so that coherent signals would arrive at the latter's brain about the bodily conditions of the former. (In fact, although this seems a possibility, I don't think we have any reason to suppose it is a very likely one.) When a fly landed on the first person's leg, the second person would feel the sensation, and perhaps slap his own leg. The second person would be perceiving that there was a fly on someone's leg, in the way that normally perceives that there is a fly on one's own leg. Because of possibilities like this, I add the qualification "normally".

Paralleling normally self-informative ways of knowing are normally self-dependent/directed/effecting ways of acting. Towards the end of the movie *Spellbound*, we see Leo G. Carroll point his gun at Ingrid Bergman as she walks out the door. Holding the gun in this way and pulling the trigger is a way, in the circumstance in which there is a person in front of the agent, of killing that

person. As the movie continues, we see Bergman continue to walk away, toward the door of Carroll's office, from his perspective. Slowly, we see the hand holding the gun turn, until the barrel of the gun is all that is visible on the screen. Then it fires. We know what Carroll has done, and to whom. He has killed someone, and the someone is him. The way Carroll held and fired the gun was a normally self-effecting way of killing someone. Of course, if Carroll had a head shaped like a donut, he could have shot the someone behind him. But normal people normally kill themselves when they shoot like that.

This is only a particularly dramatic case of a whole class of actions. Imagine George and Laura Bush seated across from each other at a boring dinner. Both know that the president is thirsty. Both may desire that he get a drink. The appropriate action for the President to take is the familiar one of reaching out and bringing the glass of water towards his lips. That is an action that will succeed if the agent is thirsty. It is a normally self-dependent/directed/effecting action. It won't do any good for the First Lady to perform it. At least, it won't help relieve the President's thirst. She should pick up the water an offer it to him. That is a way of relieving (or helping to relieve) the thirst of someone sitting across from you.

I suggest, then, that self-notions are those that have the special role of being the repository for information gained in normally self-informative ways, and the motivator or actions done in normally self-effecting ways. This does not imply that there won't be a lot of information associated with the self-notion that is gained in other ways, or that the beliefs and desired involving the notion don't motivate actions done in ways that are not normally self-effecting. Hume sent a rather favorable (anonymous) review of his *Treatise* to a journal; this was a normally other-affecting way of acting, but was motivated by his desire for literary fame *for hmself*. When he read reviews by others, he was picking up an information about other people's view of the book he himself wrote. But most of the reviews he read in journals were about other people; reading a review is not

a normally self-informative way of picking up information, just a way that occasionally provides information about oneself.

Recall that in our discussion about Grand Island and the "here-notion," we noticed that one could see it to be raining where one is and decide to take an umbrella, without knowing which city one is in. The "here"-informative nature of the perception of rain, and the "here"-effecting nature of the action of taking an umbrella, guarantee that the information is relevant to the action. Similarly, one can gain information about oneself and apply it without knowing who one is, as long as the informaton is gained in self-informative ways, and applied in self-effecting ways. A thirsty amnesiac, who doesn't know who he is, can still drink glasses of water, eat when hungry, and the like. Normally self-informative perceptions can trigger normally self-dependent actions without needing to be linked to any self-independent notion of oneself.

7. WHAT'S SPECIAL ABOUT THE SELF

There is one big difference between identity and most other epistemic/pragmatic relations, and this difference makes the self-notion virtually unique. With most of the other epistemic/pragmatic relations, a given agent will stand in the relation to different objects at different times. The place one is at, the person to whom one speaks, the food in front of one--these things change all the time. This means that one cannot use a buffer tied to one of these relations to accumulate information about a given object. But one is always identical with the same person. My self-notion can be both tied to an epistemic/pragmatic relation, and also serve as my permanent file for myself.

If one did not move from city to city, one could also use one's "here-buffer" as a permanent file for the city in which one happened to live. As a matter of fact almost no one leaves the earth, so for most of us "this planet" will always refer to it; our "this planet" buffer can serve as a permanent notion for the earth. That's why I said, "virtually unique."

Suppose one not only did not move from city to city, but did not have anything to do with other cities. One never acquired information about them, or performed actions whose success depended on them. Imagine a child, who has no knowledge that there are other cities. Such a child would not even need a here-buffer. She needs no notions of cities (or "places one lives") at all.

We often talk about the weather, as if rain and snow and sleet were states of times, rather than states of places at times. We say, "It is snowing," rather than "It is snowing here". In effect, we handle a 2-ary phenomenon with a one-place predicate. This works fine so long as we can take it for granted that we are all talking about the weather in the place where we are talking. We can also *think* about the weather in this way. So long as the place in which we pick up information about the weather and the place to which we apply that information are fixed as the same by factors outside of thought, we don't need to keep track, as the example about Grand Island showed. We can have a one-place idea for a 2-ary phenomenon (See Perry, 1986).

Consider, for example, the way we think about time before we learn about time-zones. When we look at the kitchen clock or our watches, we learn about what time it is in the time zone we are in. Looking at one's watch is a normally time-zone-one-is-in informative way of learning the time of day. If one had very long arms and very acute vision, or lived right where the time-zone boundary is, one might have to be careful. But it's pretty secure. Most of the actions we use our kitchen clocks and watches to guide, are normally time-zone-one-is-in-dependent ways of acting. This is because many of the things we do, such as getting up, eating, going to school, going to bed and the like, are things that people try to do when it is a certain time of day in the time-zone they are in. As long as our dealings with time amount to using information gained in normally time-zone-one-is-in informative ways to guide actions that are time-zone-one-is-in dependent for their success, we have no reason to even be aware of the fact that the time of day is relative to time-zones.

We need to be aware of this relativity when this condition is broken. A child may learn how things can go wrong when she speaks to her grandmother long distance, or takes a trip that crosses time-zones. Her watch still tells time accurately enough, but it is the time at home, not the time at the place the child is visiting. But it is the time in the place the child is visiting that determines when lunch is served, and when the good television programs come on, and when one is expected to go to bed.

Similar remarks apply to self-notions. For many purposes we don't need notions of ourselves at all. Consider the simple act of seeing a glass of water in front of one and drinking from it. The perceptual state corresponds to a relation between an agent and a glass of water. It is the state an agent is typically in when there is a glass of water in front of that agent. The perceptual state is then not only normally object-in-front-of-one-informative, but one-who-is-in-the-state-informative. The coordinated motion of hand, arm and lips by which the agent gets a drink is not only normally object-in-front-of-one-effecting, but also agent-who-does-the-action-effecting. The identity between the perceiver and the agent is (normally) guaranteed outside of thought, by the "architectural" relations between the eyes and arms. One need not keep track of it in thought.

Another somewhat Tractarian (Wittgenstein, 1922/1962 secs. 5.62ff; see also Moore, 1962, pp. 302-03) or Carnapian (Carnap, 1967, sect. 163) way of making this point, is to say that the world as we perceive it does not include ourselves, but has ourselves as sort of a point of origin. Suppose I tell you that one point is at (4,5) and another point is at (5,4). As long as you can assume that the points have been given, relative to the same point of origin, you know that to get from the first to the second you take one step away from the *y*-axis and one step towards the *x*-axis. But if they are given relative to different origins, you will not know what the relation between them is. Similarly, if I show you how things look through a certain pair of eyes focused on a table with a cup on it, you will know what an arm will have to do to intersect with the cup. You will, that

is, if the arm is connected to the body to which the eyes belong, in the normal way. But you would have no idea how any arbitrary arm might have to move to perform that operation on the seen cup.

The self really comes in twice over when one notes that one is hungry or that one's hands are dirty, as both the perceiver and the object perceived. And when one eats or washes one's hands, one is both the agent and the object effected. The success of sticking one's hands under the faucet, as a response to the sight of one's own dirty hands, depends on a number of identities that are usually architecturally guaranteed. When one sees dirty hands in a certain way, it is the perceiver's hands that are dirty. When one washes hands in a certain way, it is the agent's hands that get clean. And when a perception of the first sort causes an action of the second sort in a more or less direct way, the subject of the perception is the agent of the action. We don't really need a self-notion to handle any of this. We will need one when we start to get information about ourselves in ways that are not normally self-informative.

In a world like ours, with mirrors, newspapers, lists of people who are supposed to be various places and the like, we all have many ways of knowing about ourselves that are not normally self-informative. They are just the same ways we have for knowing about others. I can look at my ticket and see when I am to leave; you can look at my ticket and see when I am to leave. The ticket gives the same information in the same way to anyone who looks at it: John Perry is to leave at a certain time on a certain day. I need an objective notion of myself to pick up the information, and a self-notion to put it to use. Unless I already have or acquire of a notion of John Perry as John Perry, I won't have any place to store the information I get from the ticket. Unless this is or is linked to a self-notion, I won't end up performing the normally self-affecting actions that I need to perform (like getting out of bed) in order to get to the airport on time.

We have seen, then, that there are three kinds of knowledge concerning or about oneself. First of all, there is knowledge that doesn't require a self-notion:

knowledge that is picked up in normally self-informative ways and is not combined with other sorts of information, and guides actions performed in nomally self-effecting ways. This is the sort of self-knowledge required to drink a glass of water, or feed onself. I'll call this *agent-relative* knowledge. It is knowledge represented in agent-relative ways.

The second kind of knowledge I'll call "self-attached" knowledge. This is knowledge of oneself, however obtained, that has been added to one's self-concept or self-file by being attached to the self-notion. This is the knowledge we express with the word "I."

The third kind of knowledge about oneself really doesn't strike us as self-knowledge at all. It is knowledge a person has about himself or herself, that is not attached to the self-notion. This is the sort of knowledge that we have in the Castañeda cases. The biography writer knows that the war-hero saved many lives, but doesn't know that *he* saved many lives, even though he is the war-hero. I'll call this "knowledge merely about the person one happens to be." If we remove the "merely" we get a sort of knowledge that we all have a lot of. I know acquire a lot of knowledge about John Perry in the same way that others do; I look up the times of my classes in the *Time Schedule*, my phone number in the phone book, and forth. This is knowledge of the person I happen to be. But normally I associate this with my self-notion. When I don't, so that I am a little like Castañeda's biography-reader, it is knowledge *merely* about the person I happen to be.

8. BACK TO MACH

When he looks to the far end of the bus, Mach gets information about himself in a way that is not normally self-informative, but normally "person-I-am-looking-at" informative. This information doesn't pass into his self-notion; it is not combined with information gotten in normally self-informative ways. And it doesn't motivate normally self-dependent actions. Mach has knowledge about

Mach, but it is not attached to his self-notion; it is knowledge merely about the person he happens to be.

Suppose Mach looks down at his own vest and sees a big piece of lint. (Mach himself provides us with a picture of the way one's front characteristically looks to oneself.) This is a normally self-informative way of knowing that a person has lint on his vest. If he had seen the lint in this way, he would have associated the idea of having a large piece of lint on one's vest with his self-notion. That's what I mean by saying that the self-notion is the repository of normally self-informative perception. Now if Mach had desired not to have large pieces of lint on himself, he would have reached out and removed it in a way that works when the piece of lint is on one's own vest---sort of a downward brush with the back of the hand often works. If he has the desire to be lint-free associated with his self-notion, and the idea of having lint on the vest is associated with his self-notion, we would expect him to take such a normally self-directed and self-dependent action. That's what I mean by saying that the self-notion is the motivator of normally self-dependent/directed/effecting actions.

But when Mach sees a piece of lint on the vest of the person in the mirror he does not act in this way. The information is not gotten in the normally self-informative way. It is not combined with the other information in the self-notion, and doesn't lead to the action that works to remove lint from oneself.

At the beginning of the episode, Mach formed a notion for the person he saw, whom he took to be getting on the other end of the bus. This was a notion of himself, but not a self-notion. We assume Mach knew who he was, and so that he had a notion of Ernst Mach as having all of the well-known properties of Ernst Mach that was also a self-notion. But even if Mach had been in the middle of a bout of amnesia he would have had at least a self-buffer, a notion tied to normally self-informative action and ways of knowing. Mach's beliefs change, during the episode, in that he comes to link the new notion formed when he got

on the bus with the old self-notion or notions that he has. If, after he has made the connection, he notices that the person in the mirror has a piece of lint on his vest, he will pick the lint of his own vest in the normally self-dependent and self-effecting way of picking lint off one's vest.

9. SELF-KNOWLEDGE PROBLEMS REVISITED

Now let's return to the issues about self-knowledge that seemed to stand in the way of the simple and straightforward account of the self as the person. These peculiarities of self-knowledge can be explained by taking self-knowledge to be a species of agent-relative knowledge.

These kinds of knowledge are, like self-knowledge, "essentially indexical." We use "now" and "today" to express our knowledge of what time it is, and "here" to express our knowledge of where we are. These locutions are not reducible to names or objective descriptions, just as "I" was not. I cannot express what I say when I say, "The meeting starts right *now*" by saying "the meeting starts at *D*" for any description *D* of the present moment.

We are also immune to certain sorts of misidentification when we use certain methods of knowing. There is a way of finding out what is going on around one, namely opening one's eyes and looking (Evans, 1985). Now when one learns what is going on in this way, one can hardly fail to identify the time at which this is happening as *now* and the place as *here*. And finally, the forms of thought we express with "now" and "here" seem to have a unique motivational role. If I want to do something *here* and *now* I will simply do it.

So, to summarize. We cognize things, times and places not only objectively, but via their present relationship to us---via agent-relative roles. There are ways of knowing and acting that are tied to such roles, and our knowledge exhibits immunity to misidentification relative to such roles. And knowledge via such roles plays a special motivational role. Finally, because

different objects play these roles in our lives at different times, it is invalid to accumulate knowledge along them.

"Self" expresses an agent-relative role, that of identity. As with other agent-relative roles, there are special ways of knowing and acting that are associated with identity. If Mach had wished to know, during the interval while he was confused, if the shabby pedagogue he was seeing had lint on his vest, he would have had to walk over to him and look. If Mach had wanted to know if he himself had lint on his vest, he could have simply lowered his head and looked. Had he done this, he would have had no doubt about whom the lint was on. If Mach found lint and wanted to brush it off, he would engage in self-brushing, a quick movement of the hand across one's front that each of us can use to remove lint from our own vest and no one else's.

Unlike most of the other agent-relative roles, identity is permanent. I will have many things in front of me, talk to many people, be in many places, live through many days in the course of my life. But there is only one person I will ever be identical with, myself. I never have to unlink my self-buffer from my John Perry notion. It can be a self-notion; it can just be my self-buffer. Accumulating information in one's self-buffer for life is valid, unlike accumulating in one's here longer than one stays in one place, or in one's today-buffer for more than twenty-four hours, or in one's person-talked-to buffer longer than a conversation.

I also won't ever be on another planet, in a position to pick up information about that planet by looking around, and able to refer to the planet as "this planet" with a demonstration towards my feet. I would be relieved if this were necessary, but it is contingent. But my identity with myself is necessary. I do not claim that identity is the only necessarily stable epistemic/pragmatic relation. Perhaps it is necessary that I am in *this* universe. So identity might not be unique in providing a necessarily stable agent relative role, a buffer for which can used to accumulate information. But it's pretty special.

Earlier we rejected the straightforward account of self-knowledge as knowledge about a person by that very person. Now we can put forward an alternative. Self-knowledge is knowledge about a person by that very person, with the additional requirement that the person be cognized via the agent-relative role of identity. This agent relative role is tied to normally self-informative methods of knowing and normally self-effecting ways of acting. When these methods are employed, there will be immunity of misidentification as to whom is known about, or whom is acted upon. So agent-relative knowledge and self-attached knowledge count as self-knowledge on this definition, but we don't need to count knowledge merely as the person one happens to be as self-knowledge.

Being the person known about self-informative ways, and the person affected by actions done in self-effecting ways, can serve as a person's fundamental concept of himself or herself. In this way our self-conceptions have a different structure than our conceptions of other individuals of importance to us. If we understand the special way in which a person's self-knowledge is structured, we do not need to postulate anything but the person himself or herself for the knowledge to be about. [iii]

---

[i] I am ignoring the complication provided by other people named "Ernst Mach".

[ii] I am ignoring the complication that "here" isn't usually a name or pronoun referring to a place but an adverb of place.

[iii] Various versions of the material on which this paper is based have been given in lectures at Stanford University, Notre Dame University, Cornell University, the University of Wisconsin at Madison, the University of California at Davis, Princeton University, and Santa Clara University. I am grateful for the many helpful comments that have been made on each of these occasions; I'd particularly like to thank David Copp and Carol White.