Selves and Self-Concepts[1]

John Perry

Stanford University

1. Introduction

Some philosophers think of selves as rather mysterious things. Sometimes

selves are identified with the souls of Christian theology, or the essential

natures that are passed along in reincarnation, or as noumenal objects that

exist beyond normal space and time, outside of the causal realm, and join,

in some Kantian way, with the primordial structure of reality to create the

world as we know it. And Hume famously could find nothing to serve as

his self, except a bundle of perceptions.

I am inclined to think that most of these deep thoughts are not

required to understand the self. I am inclined to think that a self is just a

person, thought of under the relation of identity. I am also inclined to

think, however, that Hume had something importantly right with his bundle theory. In this paper, I will explore both ideas.

2. The Straightforward Theory

Consider what it is to be a neighbor. A neighbor is just a person, thought of as having the relation of *living next* to some person in question. A teacher is just a person, thought of as having the relation of *teaches* to some student. A father is just a person, thought of under the relation of *father of*. People play important roles in other people's lives, and we give these roles titles: neighbor, teacher, father, spouse, boss, and so forth.

But we each play an important role in our own life. I have a relation to myself that I don't have to anyone else, identity. Self is to identity, as neighbor is to *living next door to*. It is a way we think of ourselves. The self is not of a special kind of object, but the concept of self is a special kind of concept, that we each have of ourselves as ourselves. It is the unusual structure and special causal and informational role of our self-concepts that underlies the philosophical problems about the self.

2.1 Castañeda's War Hero

Sometimes the best way to find something is to first consider a case where it stands out because it is abnormal. Now a sort of paradigm case of someone who doesn't know who they are, and in that sense lacks an ordinary sense of identity, and has a diminished self-concept, is someone who has amnesia. Here I am thinking of a certain kind of amnesia, which may only exist, in its most perfect and full-blown state, in fiction and in philosophical examples. This is a person who, as a result of a bump on the head, has no idea who they are. One assumes that the knowledge is somewhat still in the brain, waiting to be released by another fortuitous bump on the head, or perhaps surgery, or perhaps simply time. Such people do not lose their "know-how". They can still use language, feed themselves, walk, and the like.

I'll use an example from the great late philosopher Hector-Neri Castañeda (1966). He imagines a soldier--call him Bill--who having performed many brave deeds in a great battle, is injured, loses his dog-tags, wanders far from the scene of his heroics, falls into a coma, and awakens with amnesia. Not only does he not know who he is, no one else

does either. He is clearly a solider, however, and clearly due all the rights

pertaining thereto, so he is hospitalized, cured of everything but his

amnesia, and goes to Berkeley on the GI Bill. In the meantime, Bill's feats

during the battle have become well-known. People don't know what

became of him and assume died in battle, as this seemed inevitable given

the situation he was in when last seen. He is awarded many medals

posthumously.

For the time being let's concentrate on Bill, lying in the hospital, just

awakened from his coma, before he is even told how he was found at the

scene of the great battle. Now of course there is a sense in which he does

know who he is. He can say, "I am me." Suppose Bill feels a pang of

hunger, and sees a piece of chocolate cake on the tray in front of him. Does

he wonder into whose mouth this morsel should be put, in order to

relieve his pang of hunger? No. He knows that *he* is the person who is

feeling the pang of hunger, and the person whose arm he can control more

or less at will, and the person who has a mouth which he can't see right

below the nose the tip of which he can see, and he knows how to direct

the fork and the cake into that mouth. He knows that he is sitting in a

room on a bed, with a window out onto a lawn, maybe with a radio and some magazines on the stand beside him. So, he really knows a great deal about himself. Still, compared to the rest of us, he has a very diminished sense of self. He doesn't have memories from which he can construct a narrative about why he is where he is. He doesn't know what values, what commitments, what beliefs, what actions led him to this hospital room.

Since he doesn't know his own name, he can't exploit public sources of information about himself. He picks up the edition of *Stars and Stripes* on table beside his bed, and reads all about the hero who has just been posthumously awarded many medals. He does not feel pride in, but only admiration for, all of the brave things which he, in fact, did. At this point he has two concepts of the person he has to be, his concept of the war hero, and his self-concept.

Normally, we rely a great deal on such public sources of information about ourselves. If I forget my phone number, I look it up in the Directory. I find out something about myself in exactly the same way as you would find out the same fact about me. Indeed, there are lots of things that make

it into the public conception of us, that we can't discover in any other way. To find out the time of my talk at this conference, I looked at the program on the net. I found out when John Perry's talk was scheduled the same way everyone else did.

In contrast, all of the knowledge Bill has about himself in the hospital (or almost all), he acquires by what I will call, somewhat ponderously, "normally self-informative ways of knowing about a person". That is, when you see an object by holding your head erect and opening your eyes, the object you see will be in front of someone. Who? You. Normally, at least, this is a way of finding out what is going on in front of the person who is doing the seeing. You feel a pang of hunger: someone is hungry, and will have their hunger relieved if food enters their mouth and makes it to their stomach. Who? You. This information is, in Shoemaker's words, immune to error through misidentification.[1] If you think you see a tree, you may be wrong. Perhaps it is a large bush, or a skinny man. But you won't be wrong about whom the thing you are wrong about is in front of. It will be in front of you. Why do I say "normally self-informative"? Perhaps some day brain scientists will invent a little device that will send

message from one person's eyes to another person's optic nerves, so that the second person can directly see what is front of the first. This might have some military utility. Old, frail, jittery, demolition experts can guide the movements of young, healthy, steady, inexperienced ones, as they defuse bombs. These experts will then have a cognitive burden that is not placed on most of us. They will have to keep track of whose environment they are visually obtaining information about. Most of us don't have to do that. They will have lost immunity with respect to one way of getting information. Technology can do that.

Some ways of knowing about ourselves are not only normally but necessarily self-informative, or so it seems. Our veteran demolition expert, if he gets confused, may make a mistake and think the pompous general in front of him is really in front of the novice. He may say something he shouldn't about pompous generals. But could he be wrong about who has a headache that he feels? Well, no sooner do we admit that something seems necessary than some cleverer philosopher will come up with a scenario that shows that it isn't.

Now consider Bill's act of extending his arm, grabbing his fork, breaking off a piece of cake, and shoving it in his mouth. I'll call that a "normally self-effecting way of acting". Moving in that way is a way anyone can shove a piece of cake they see in front of them in their own mouths, a way of feeding themselves. Again, normally, because we can dream up cases where it wouldn't work.

I'll repeat my favorite example here. At the end of Alfred Hitchcock's movie "Spellbound" Leo G. Carroll holds a gun pointed at Ingrid Bergman, who is leaving his office, having just exposed his plot to frame his patient, Gregory Peck, for murder. We know who Carroll will shoot if he pulls the trigger: the person in front of him. Shooting a gun pointed like that is a way of shooting the person in front of you. Then we see Carroll's hand turn the gun around. The front of the gun barrel fills the whole screen. He fires. Whom does he shoot? Himself. Firing a gun held like that is a normally self-shooting way of acting. But suppose that Carroll had a donut-shaped head. Then it would be a way of shooting the person behind him. It's only a contingent fact that we don't have donut shaped heads. That's why we need to say "normally."

Note, however, that even if he had not turned the gun around, and had gone ahead and shot Bergman, Carroll would have performed some normally self-effecting actions. To kill Bergman he fires the gun, and to fire the gun, he pulls the trigger. He brings it about that his trigger finger moves, in the normal way we do such things, but just doing them. In this sense, every action begins with an *executable* action, a movement of the agent's body or some part of it, a movement the effects of which will, if things go right, lead to the goal for which the action was done.

So Bill, even with his amnesia, has a good deal of self-knowledge, in a perfectly reasonable sense.

Bill proceeds to Berkeley, where he ends up getting a graduate degree in history, writing, for his dissertation, a biography of the war hero who gained his fame at the very same battle from which Bill woke up with amnesia. He doesn't figure out for quite a while--after he's earned his Ph.D.-- that he must be the war-hero, that his dissertation is actually autobiography. Now the point of this is that Bill knows a great deal about a person, who happens to be him. In a sense, he knows a great deal about himself, for he knows a great deal about a certain person X, and he is X.

But that's not what we would ordinarily say. We would say something

like this: Bill knows a great deal about the person he happens to be, but he

doesn't know much about himself. We have to distinguish between self-

knowledge proper, and knowledge that happens to be about oneself,

knowledge of the person one happens to be.

2.2 First Person Memory

In fact, even when Bill finally figures out that it is him he is writing about,

we might be reluctant to call what he is writing an autobiography. One

important thing John Locke[2] and Shoemaker emphasize is that we have a

special access to our own past thoughts and actions. We remember them--

but we can remember the *past* thoughts and actions of others, too. I can

remember that Elwood used to think that poison oak was edible; I can

remember the time Elwood ate some poison oak.

But in the case of my own thought and action, I not only remember

that someone did something, or that someone thought something. I

remember thinking and doing things. Shoemaker calls this "first-person

memory" and "remembering from the inside". Our access to our own past

thoughts and actions is phenomenologically and logically different than our memories about what others have thought and done. Remembering doing something is not *like* remembering someone else doing something. And in the case of others, there is always the questions of *who*? I remember that someone eating poison oak, but was it Elwood? But if I remember *eating* poison oak, it was me that was doing the eating. Maybe it wasn't poison oak but poison ivy or just some cilantro I had an allergy to. But at least it was me that did the eating. Immunity carries over into first-person memory.

Of course, some combination of technology and imagination might undermine this too. Perhaps between the eating and the remembering the person who did the eating, the person I used to be, split into two people, neither of which can exactly claim to be the very person that did the eating. Then I only quasi-remember (Shoemaker's term again) the eating. Or suppose that memories were stored in the blood. Suppose Bill Clinton donated to a blood bank and I was given some of his blood. All of a sudden I am having flashbacks to afternoons spent with Monica Lewinsky and thoughts of Paula Jones. Let's not go further in this direction. (I have a

feeling the memories-in-the-blood idea also came from Shoemaker's fertile imagination, but not the Clinton example.)

Once Bill figures out that he is the war hero, he can assimilate all the facts he has learned about his own to past into his own self-concept. But he still won't be related to these things in the normal way, the way we expect of an autobiographer. He will know that he did these things, and after a while he will remember that he did these things, but he won't remembering doing them.

A similar distinction applies to our knowledge of what we will do in the future. I can know, or at least have a pretty well-grounded belief, what you intend to do, and what you will do. But when I know what I am doing, what I am trying to do, what I intend to do, and in those ways, what I will do, it is based on a different way of knowing, a way each of us knows something of his own future; again, it is knowledge from the inside.

A case like Bill's is pretty fantastic, but the underlying moral is generally applicable. It is a fact about the complex informational world we

live in that we have lots of ways of getting information about ourselves that are not normally self-informative.

Self-knowledge, in the ordinary sense, is knowledge of ourselves that is part of our self-concept. Knowing facts about the person you happen to be, as Bill did when he wrote his dissertation, isn't enough.

3. Knowledge from a Perspective

We really need to recognize a third kind of self-knowledge to complete the picture. Any animal picks up information from a perspective and acts from a perspective. Its only way of picking up information will be in normally self-informative ways. It has no real need of a self-notion to keep track of facts about itself. Except for occasionally stumbling across a still pond or a mirror, it doesn't pick up information about itself in normally other-informative ways. This primitive sort of self-knowledge I'll call *knowledge from one's own perspective.*

Now we might say that such being has no self-knowledge. I'd rather say it has a primitive self-knowledge. The animal itself is what I call an

unarticulated constituent of its whole system of knowledge of the world

around it.

Here is what I mean by this. I think of propositions as abstract

objects that we use to *classify* states and events that have contents,

paradigmatically thoughts and utterances. In the case of beliefs and

statements, the propositions capture truth-conditions. Often we expect a

isomorphism of sorts between the whatever mental or linguistic

representations are involved and the propositions we use to classify the

states. If you say, "Brutus stabbed Caesar," your statement has the content

that Brutus stabbed Caesar. The constituents of the propositions, Brutus,

Caesar and stabbing, correspond to words in the sentence you used. The

belief that you were expressing involved your concepts of Brutus and

Caesar and stabbing.

Sometimes we just leave out words when we do have the concepts. If

you ask me the time, I'll look at my watch and say "It's two." If my wife

asks me the time, when we are talking long-distance, I might say "It's two

o'clock here" or "it's two o'clock eastern time." I am aware that o'clock

attributes are relations between times and places and not just properties of

times, but I don't bother to note that relativity when its obvious to what place my report is relative. When it isn't obvious, as in the conversation with my wife, I do explicitly note it.

But sometimes the constituent is not simply unexpressed but also unthought. I ask Jamaica-Unique, my 7 year-old granddaughter, what time it is. She looks at the clock and says, "It's 11 o'clock." "That's right," I say. What makes her statement true is the fact that it was 11 o'clock Pacific Standard Time. The time-zone is the unarticulated constituent.

It might be misleading to say that she *expressed* the proposition that it was 11 o'clock Pacific Standard Time , or that that is what she said. But that fact is what made her statement true. If I was talking to Jamaica over the phone from New York, perhaps really needing to know the time since my watch has stopped, I would learn that it is 2 o'clock Eastern Standard Time. It is my knowledge of the relation between the time zone that makes her statement true and the time zone I am in that makes this inference possible. Where a constituent is unarticulated in both speech and thought, I'll say that the thought "concerns" it, as opposed to being

straightforwardly about it. Jamaica's thought that it was 2 o'clock

concerned California.

Classifying the truth-conditions of beliefs and utterances with

propositions that are not articulated in those beliefs and utterances is often

an important part of explaining how those beliefs and utterances work to

provide the agent with information, or reflect the information they have.

Jamaica has the ability to look at an accurate clock set for the time zone in

which it resides, and report accurately what time it is in that time zone.

She has a way of getting knowledge about what time it is, a method that is

normally informative about the time-zone one is in. If Jamaica had

extraordinarily acute vision, or lived in a house on the border of two time

zones, this ability would be undercut--like that demolition expert we

considered earlier, she would have a cognitive burden on her use of clocks

to tell the time that most of us don't have.

Just as Jamaica doesn't need to keep track of the time-zone relative to

which she is telling the time, animals don't have to keep track of which

animal's environment they are getting information about. It's always their

own. It seems that the cognitive, informational needs of many animals

would not gain much by having the sort of conceptual architecture we have for dealing with self-knowledge.

Such an animal has primitive self-knowledge, and we can, if we want, say that its whole cognitive system, everything it perceives and believes and wants, constitutes its self-concept. Such a self-concept wouldn't be a component of thoughts, however. Such an animal wouldn't have a self-concept part of which was a concept of the animal it happened to be, similar, to its concept of other things. A simple animal will think of everything via the role that thing plays in its present experience: object in front, object to be eaten, and so on. A more sophisticated animal may have detached concepts of particular things; certain other animals and things that it can recognize, and respond to in terms not only of its presently manifested properties but in terms of properties the animal has perceived them to have in the past.

What many animals don't have is a concept of themselves, that is like these other detached or detachable concepts it has. It doesn't take itself to be one of the animals among others in the way we do. We do so by having a concept of the person we happen to be, linked to our

primitive self-concept, to normally self-informative methods of knowing and normally self-effecting ways of acting. I would not want to claim that language is a necessary condition for needing such a concept. But it is clearly involved in the necessity for our having such a concept. For as we noted, we get information about ourselves not only in normally self-informative ways, but also in normally other-informative ways, largely in virtue of language. If we had a developed a protocol language to express primitive self-knowledge, it would have sentences like "Tree in front" and "Move paw towards food in front". Once we have a self-notion, we can perform inferences into and out of that language that amount to "I-introduction" and "I-elimination". From "Tree in front" we infer "Tree in front of me" and from "Tree in front of me" we infer "Tree in front". This latter sort of inference suggests how information from our self-notion gets translated into action, even for sophisticated animals like ourselves with self-concepts. The self-concept is the concept formed by connected to this more primitive kind of self-knowledge a concept of the same type as we have of others.

Now if we simply put in our explicit self-concept, through I-addition, the things we gain from self-informative perception and memory, and then take them back into our primitive self-concept, through I-elimination, it would seem a pointless exercise. But because we have other ways of knowing about ourselves, it is not. I learn, using a normally other-informative method, that John Perry is supposed to lecture at noon. I know that I am John Perry, so I can infer that I am to lecture at noon. Then, by I-elimination, I get: Lecture at noon. The opposite process allows me to make information, gained in normally self-informative ways, available to others. I feel an urge I might primitively express as "Want a beer now". I-introduction is helpful in a bar: "I want a beer." For room service in a hotel, that won't do. I have to say, "John Perry, room so and so, wants a beer." Any architecture that allows us to get to our lectures on time and order a beer from room service clearly has a lot going for it.

All actions, however distant their intended results extend, begin with movements of our own bodies and their parts, intended to have effects on objects playing roles in our lives at the moment of action. I want to send a fax to Norway; the result I have in mind is getting Dagfinn Føllesdal to fax

me back a draft of an article he is writing. It is getting this request in front

of Dagfinn that is the motivating intention, the one thing I want to

accomplish. I could do it by calling, or shouting if I had a loud enough

voice, or smoke signals, or email; faxing is just one way to bring about the

result. So this intended result doesn't dictate much in the way of bodily

movement. But once I have built my plan down to the point where it is

executable, I need to move my body to get it started. I need to stick my note

in the fax machine and poke the right button with my finger.

It is only by being attached with our self-concept, and hence with

knowledge from a perspective, that our beliefs and knowledge about the

world at large is of any practical use to us, as forming the basis for

initiating action. I desperately need Dagfinn's paper. I know the paper

exists, that Dagfinn is authoritative, that I will be able to understand what

he says about Husserl, and that he will be willing to let me see the draft. I

carry this information about Dagfinn around in my head even when I am

not with him, or on the phone with him, even when I have no knowledge

where this inveterate world traveler might be. I don't think of him via a

role he plays in my life. Perhaps that is overstating it, but at any rate no

role that leads back to executable action. There isn't much I can do. But if I know he is in his office in Norway, I can send him a fax. I can think of him as a person who will read the fax that I send by typing certain numbers into the fax machine in front of me. He is attached, however tenuously, to my self-concept, he is the person who is playing a certain role in my life that provides the possibility of interaction, as the person who will read the fax I will send by moving my fingers in a certain way.

This structure of our self-concepts, linking concepts about the person we happen to be (or at least take ourselves to be), with our primitive self concept and hence with what we perceive and what we do, is very basic to human life. We incorporate what others notice and know about us into our own self-conception. We do this all the time. And in fact most of us are very concerned about our *public identities*: the conceptions of us that others have, what our mothers and fathers and sons and daughters and colleagues and students think of us. It is what is written next to our names in the newspaper or the college catalog, or on the vita on our web page. It is what presidents of the United States worry about when they worry about their place in history. For many issues, it is a better source of

information about ourselves than any normally-self-informative method of knowing.

For most of us some very important building blocks of our own identity, our own self-conception, come from the outside, from assimilation into the "I" of the "me"; that is, by adopting as part of our self-concept opinions about ourselves that originated with the insights, or mistakes, of others. My parents tell me that I am like my grandfather, that I am a thinker not a doer, and that becomes part of my self-conception. Through a series of accidents I become a philosopher professor, and slowly the expectations the world has of philosophy professors seeps its way into the core of my self-conception. I quit worrying about attributes that once seemed to be faults: flakiness, irreverence, inappropriate use of humor, procrastination, difficulty with the more practical aspects of life. I begin to take pride in them.

Public identities may take on a life of their own. Thus Borges:

I know of Borges from the mail and see his name on a list of professors or in a biographical dictionary. ... It would be an

exaggeration to say that ours is a hostile relationship; I live,

let myself go on living, so that Borges may contrive his

literature, and this literature justifies me.... my life is a fight

and I lose everything and everything belongs to oblivion, or

to him ("Borges and I," 246-47).

We may have several public identities. I may be one person in the eyes of

my surviving cousins, who meet every so often in Nebraska and reminisce

about our grandmother and grandfather, uncles and aunts and each other.

My interest in Nebraska football, rainfall, and whooping cranes comes to

the surface. It's not an act; it's just an motivating complex coming off the

shelf, in response to the situation. It goes back on the shelf when I get back

to Stanford.

So I have a sense of my own identity. Here we see another use of the

term "identity," perhaps more ordinary than the one philosopher's use.

What is my identity? For some psychologists it is the most central part of

my self-concept, the things I think are not only true of me, but in some

sense define me. Perhaps I can imagine myself as being a woman, or a

Canadian citizen, but I can't imagine not being a philosopher; it is unthinkable. Then being a philosopher is part of my identity in this sense.

4. Notions and Bundles

As I conceive of beliefs, desires and concepts, they are not abstract contents, not properties, not Fregean senses, not propositions, but cognitive structures in particular minds that have causes and effects. Concepts *have* contents; that is, we classify them by things they are of or about. A concept is not a concept *of* a certain thing because that thing fits its content. It is a concept of a thing because that is the thing it is used to hold information about, and guide behavior towards. To make sense of this picture, we need to recognize *notions* of particulars and *ideas* of properties and relations.

I think of notions as file folders, into which information of various sorts is put, and I often explain them in this way. A student comes into my office, who has been assigned as my advisee. I take out a file folder and dedicate it to the job of storing information about this student. I put the various documents that come to me about the student, the work he gives

me, my notes concerning his progress or lack thereof towards a degree. The whole thing is like my concept of the student; the file folder is what makes it of *him*--even if, in my careless and slothful way, I get things misfiled, and the folder is full of misinformation about him.

The stuff that's put in the file folder provides the "what" of my concept of the student, the file-folder the "who". Together they constitute what I believe about him and what if anything I desire for him.

Now consider our friend Bill, working away at his dissertation on the war hero. He has two concepts of Bill. One is a concept of Bill because it is the concept assigned to keep track of a certain person, the one Bill read about in *Stars and Stripes*. His other concept of Bill, his self-concept, is a concept of Bill because it is Bill whom it stores normally self-informative information about, and because it is Bill's executable and other self-effecting actions that it motivates. The more primitive part of his self-concept was created to keep track of Bill only in the sense that his whole cognitive system was created, along with the rest of him, with the function of keeping track of what was going on in the world from his perspective, and initiating actions that made sense given this information. The other

part of his self-concept began its life when Bill was in the hospital, with amnesia, and had to begin anew building up a concept of the person he happened to be, to link with his primitive self-concept, and to provide a place to put information he got in other ways--what the nurse told him about where he was found, and so forth.

Let's push the file-folder metaphor a bit further. Think of Bill's mind as a filing cabinet on wheels. It is full of file folders, for the persons, places and things that Bill has encountered, read about, and otherwise come to be acquainted with, however, indirectly, to the extent that it has been worth-while to form more or less permanent concepts of them. These file folders are notions; they have histories; they were introduced when Bill met, read about, saw, or otherwise encountered objects worth keeping track of. Lots of these files will include information that will allow Bill to recognize the objects these files are of, should he encounter them again. Sometimes people he hasn't encountered, like General Smith, include information that will allow Bill to recognize them when he sees them for the first time. General Smith, according to the file, is a big ugly fellow with puce colored hair and lots of medals.

On top of the filing cabinet are two boxes, the inbox and the outbox. All the notes in these boxes are in protocol language. The inbox has entries like "Mild headache, now" and "big ugly fellow in uniform with medals and puce colored hair in front." The outbox has "walk forward and extend hand," and "utter 'I've always wanted to meet you General Smith'." The first file in the top drawer has entries that correspond to those in the inboxes and outboxes, like: "I have a headache," "There is a big ugly fellow in uniform with medals and puce colored hair in front of me." This file is Bill's self-notion, and the file plus what is in it is his self-concept. Included will be lots of information about Bill--what he has come to believe about himself since awakening from his coma--that he was once a soldier, that he has amnesia with respect to events before coming out of the coma, that he is a Ph.D. student at Berkeley, and so forth.

By the time he is well into his dissertation, Bill will have a pretty big file on the war hero. So he has two notions of himself. One is connected causally to him, through the story he read in *Stars and Stripes*. Another is of him, because it is connected by I-addition and I-deletion to

his input and output buffers. When Bill finally figures out who he is, he merge his war-hero folder and self-folders.

The file-folder metaphor has its limits. Mark Crimmins (1992) develops some other metaphors or models of concepts that are better for many purposes. Our notions and ideas seem to provide us with something more like a relational data base, a richer informational structure than file folders provide.

But those issues don't bother me too much today, so I'll stick with the file folder metaphor, at least long enough to get clear about another aspect of it that seems misleading.

File folders are very passive things. We do things with file folders; open them, read them, perhaps plan our actions around the information in them. But who would the "I" be that read the file folder that constituted his self-notion?

On the straightforward theory, the self is the person, the agent, the subject, the flesh and blood human being, or something very close the the flesh and blood human being, differing only perhaps in the way we would individuate it through time in certain puzzling cases that the great

philosophers of personal identity, especially Locke and Shoemaker, have induced us to worry about. Does the person then consult his own self concept? Sometimes. I might search my mind for a memory of a trip that I know I took as a kid. But the person who consults his own self-concept is already someone with beliefs and desires. I believe I took the trip and I want to know more about it. Usually the beliefs and desires in my self concept just effect what I do; I don't have to search for them.

If we think of our self-concept as a big relational data base, then we can also suppose that somewhere in our cognitive apparatus is a sort of processor with the ability to scan data and do something with what it finds, maybe several of them, busily scanning files and rewriting them in response to what they find and new input from sensors. But the processors works the same way in me as they do in anyone else. Its operations are not governed by my beliefs and desires, but rather their operations are part of what is involved in those beliefs and desires being more than mere data, but being the basis of action. The processor is not me; it doesn't do what I do, although its working the way it does is part of me doing what I do. It is part of my cognitive apparatus, but it is not me.

It seems better to think of the self-concept, the self-notion and its associated ideas as what I'll call a motivating complex. A motivating complex has two sides. On the one hand, the cognitions themselves cause movements. On the other, the contents of the cognitions provide one sort of reason for the actions that the movements constitute. Such cognitive complexes are physical, active networks in the brain, that effect the body to move in ways that will promote the desires that are part of the complex if the beliefs that are part of it are true.

Now, when we think of things in this way, Hume's bundle theory, or something like it seen through modern eyes, seems attractive. Even if my self is just boring old me, thinking about myself as myself brings in my self-concept, a network of ideas associated with my self-notion. But what is this self-notion? It might seem that buried somewhere in my self-concept must be some characteristics that *make* the folder a folder of *me*. My self-notion plays the role, semantically, of a rigid designator or directly referential term. The thoughts and beliefs memories and anticipations and fantasies I have with the self-notion as a component are thought and beliefs and fantasies about me. But what makes them about

me? Not any information about myself that is so intimate and essential to me that it finds me in any possible world no matter how different I might be there than how I am now. What makes this file a self-notion is its connection with my primitive self-notion, with the inbox and outbox in terms of the file cabinet analogy.

One idea that I think we should resist is that it is the presence of the word "I" in the file, or in the expression of my self-thoughts, as uttered to others or to myself, that makes the file a file of me. The word "I", like all words, gives us a way of doing something. "I" gives us a way of referring to ourselves. Saying "I", like moving ones hand to the top of ones head and moving ones fingers, scratching one's head, is a normally self-directed way of acting. The one is a way of scratching our heads, the other is a way of referring to ourselves. A person who has a notion that is related in the right way to self-informative perceptions and self-directed actions will use the word "I" to express beliefs about himself or herself. And this way of using "I" is what gives it its meaning, or sustains it, once it has it.

Just the bundle itself, or maybe the baling wire used in bundling it, or the file folder, nothing more transcendent, more pure, less bound by the confines of space and time and causation,than that.

Suppose that my office at Stanford were actually quite a bit better organized than it is. The university has hired efficiency experts to set up Professor's offices, as the problem of absent-mindedness became serious. All the information relevant to all of my tasks is carefully organized into file folders stored away in file cabinets in an orderly way. There is also a special set of file folders in a bin on the desk. There is no name on the bin. In the first bin is a section for all of the tasks I have agreed to to perform, and section with my long range goals and shorter range intentions and desires of the day, and also a section with my vita and some other personal information of importance. There is a huge John Perry file in the cabinet with all the other files, and a bright red ribbon between it and the bin on the desk. That's to remind me who I am. In terms of the basic architecture, that's the only real difference between my office and, say, my colleague Michael Bratman's; in his the red ribbon goes from the Bratman file to the bin on his desk.

In the morning I go in, look over the bin on the desk, and start working through the tasks. I am like the character in the movie *Memento*[3], who has lost his ability to consolidate short term memory. Trying to solve a murder in this condition, he must write stuff all over his body so that when he wakes up he will know what he has discovered so far and what he must do next. This is quite unlike the relationship I have with the data stored in my head, with my self-concept and with all of the other pieces of information with which it is connected.

Now suppose that a new kind of manila folder and a new kind of filing cabinet and some new office equipment are invented, and they are all wired together or perhaps just wirelessly communicate with one another. They are integrated into a system called the "Virtual Professor." I am no longer needed at all. The little red ribbon between the bin on the desk and the John Perry file is replaced with a red cable. The university is willing to invest in Virtual Professor systems because efficiency experts tell them the system will work better than the present one involving real people, and besides of it saves providing parking. My Virtual Professor doesn't quite claim to be me; it claims only to be VPJP, the Virtual

Professor replacement for John Perry. It takes over all my students and committee assignments and the like. The information about me, that comes via the red cable, is presented as if it is information about VPJP, with an occasional disclaimer, "For certain purposes, as part of the university's commitment to keep tuition increases as low as possible, and provide students better parking, I pretend to be John Perry, the professor I replaced."

Every morning at 8 or so the office just comes alive. The first task in the task bin lights up. A process of automated reasoning begins. The goal of the first task and the related desires and values starts feeding information into a deliberation buffer, automatically pulling up information from all the various files, including my own, as it seems to be needed. Maybe it checks the internet or makes phone calls when information is needed that is missing. Maybe it googles "John Perry - Barlow" every morning in case I have become famous.[4] Little video cameras feed in information about the students who come to office hours; they are recognized, and responded to appropriately. The deliberation buffer sends a message to the volition buffer, which starts the various

machines working. Papers are graded, emails are sent, utterances made to the students, and so forth. And, importantly, in response to phone calls, emails, consultations with the calendar, and the like, new goals are formulated, prioritized, and put in the bin. The office just churns away until five thirty. Then the office pours itself a big glass of Irish whiskey and takes itself offline until 8 the next morning. My dynamic office isn't just a bundle of things, but a dynamic bundle of things, that changes in response to information and affects the rest of the world in various ways.

I am unlike the Virtual Professor in a number of ways, including having sensations and being conscious of having them, and having a somewhat richer emotional life, and of course being much less organized and single minded, and needing parking. Still, it seems to me that the way my cognitions work is very much like a Virtual Professor's. My beliefs and desires (using those terms very generally, for all kinds of doxastic attitudes and pro-attitudes) combine to produce various executable actions. My self concept is an active bundle of cognitions that converts information and misinformation of various sorts into action.

5. Two Pictures of the Self

I will end by discussing one very important way that I, and I think most people, are quite different from the Virtual Office. This difference contributes to our sense that the real self, the real "I" is somehow elusive.

I want to contrast two pictures of our motivating self-concepts. One seems to be held by many philosophers and perhaps also by economists. Some of the people who hold it may instantiate it for all I know. I call this the *rational agent* picture. The other seems more accurate to me; I officially call this the "bundle of cognitive complexes" picture, but for short I call it the *impulsive self*.

My Virtual Professor, and Mr. Spock of *Star Trek* seem to be rational agents. When Mr. Spock is faced with a decision, he deliberates taking into account all of the goals he has and all that he believes. His desires are ordered by their importance; his beliefs by his degree of confidence in them, and that degree of confidence corresponds to the evidence he has for them. He rationally computes what the best thing to do is, that is, the thing which has the optimal chance of promoting his most important goals, given the beliefs in which he is most confident.

In other words, his self-concept forms a single, ordered, cognitive complex, and this complex motivates his actions. Emotions of course play no part in Mr. Spock's decisions, because he has no emotions; if he did happen to have one, I suppose he would wait for it to subside, or just ignore it. My conception of the impulsive self is based mainly on me, but I think it fits almost everyone I know somewhat better than the rational agent picture. My goals and beliefs combine into clusters, often with many common elements, that vie for control of my various systems of effectors. Victory is seldom complete.

For example, I'm at a department meeting. Someone says something which I interpret as a put down. I become angry. But part of me, a little inner voice, urges restraint. "Calm down. You can be almost certain you won't lose anything by shutting up. Etc. Etc." This part of me continues thinking, even as I am uttering angry words. It has no control of an important part of my agency, that part of it that controls speech. Another cognitive cluster is in charge of that.

Some goals that are quite important to me, like not appearing foolish, not alienating colleagues, not saying things that will be

counterproductive to the deliberations of the department, play no role in this second cluster, that has seized control of my mouth. They are present; they are motivating the ineffective little voice, but can't gain control of anything else, in particular they don't control what I am saying. The goals in control are ones that are not very important to me, or so I would have thought, things like making sure people know how I feel, defending myself from criticism, and the like.

What we seem to have here is two different complexes of desires, beliefs, habits, etc., each associated with the same self-notion, each parts of the same self-concept, producing different actions with different goals at the same time. The one is producing verbal behavior; the other is simultaneously producing articulate and reasonable inner thoughts and trying to take over the verbiage. Where am I in all of this? Looking back on it, I may feel like yet a third center of motivation, adjudicating between the two. But at the time, there are just the two centers.

Another familiar case, at least to me, is procrastination. One center of agency is busily making decisions about what to watch on TV, and controlling the remote, and the rest of the body, sprawled out on the

couch. The other part controls a corner of active thought, saying "Papers to grade, Papers to grade". Maybe, at some point, for some unknown reason, I move towards my study with its deskful of papers. The one complex of desires and beliefs has managed to get control of my body, the other is limited to plaintive questions that pop up unbidden in the imagination: How did Elaine get out of the jam? Did Steinbrenner fire George? And Kramer, what did he do about the car he wrecked?

It would not be correct to say that these centers are multiple selves. There is a constant activity of trying to maintain coherence and order. Most of us, most of the time, keep some kind of order among our competing centers of agency. But we shouldn't think of ourselves as somehow separate from them, adjudicating between them. The competition may produce a third center of agency, a new coalition of beliefs and desires, led by the goal of producing inner equilibrium. But this new center is not the real me, but one more part of me vying for control.

Emotions are feelings that often divert us from single-minded rationality. It seems likely that emotions often serve to in effect divert us

from deliberation, and move us more quickly to action. They are perhaps ways we tell ourselves that the case is too clear, or the time too pressing, for deliberation: act. Sometimes this is helpful. He who hesitates is lost. Sometimes it is not helpful. Fools rush in. At other times emotions empower one complex and weaken another. Sometimes they are like the elusive momentum in a football game, shifting inexplicably from one team to another. Guilt in particular pops up unexpectedly, too late to stop the sin, but in time to ruin full enjoyment of it.

6. Conclusion

Some philosophers will think the straightforward account of the self insufficiently mysterious or profound. Some will find my picture of self-concepts as bundles of competing bundles as at best a good description of a certain kind of neurotic, whose inner life falls far short of some rational ideal of personhood. I rather enjoy just being me, and take great pride and comfort in the occasional bursts of rationality that I see in myself and other humans.

Notes

1. A number of Shoemaker's important papers on personal identity can be found in Shoemaker 1984.

2. In his *Essay Concerning Human Understanding,* 2nd edition, Book ii, Chapter 27.

3. Christopher Nolan, director; see http://www.imdb.com/title/tt0209144/.

4. But not as famous as John Perry Barlow, hence it's best to add the "-Barlow" when searching.

References

Borges, J. 1964. "Borges and I" in Labyrinths: Selected Stories and Other Writings, New York: New Directions.

Castañeda, H. 1966. "'He': A Study in the Logic of Self-Consciousness." *Ratio* 8: 130-157.

Crimmins, M. 1992. *Talk About Beliefs*. Cambridge, MA: MIT Press.

Locke, J. 1972. An Essay Concerning Human Understanding, ed. Peter Nidditch. Oxford: Clarendon Press.

Shoemaker, S. 1984. *Identity, Cause and Mind*. Cambridge: Cambridge

University Press.