

# Diminished and Fractured Selves

John Perry

From: *Personal Identity and Fractured Selves*. Debra J. H. Mathews, Hilary Bok and Peter V. Rabins (Eds). Baltimore: The Johns Hopkins University Press; 2009. 129-164.

## 1 Introduction

A large part of our daily life is based on knowing what to expect from human beings, and we are amazingly good at it, given how complicated humans are. After all, my brain and your brain are about as complicated as anything the world has to offer, and they control rather large systems, hundreds of pounds in my case, that to a casual observer might appear completely unpredictable. Consider all of the people in this room. If we traced our trajectories back over the past few weeks, we would probably find paths that went all over the U.S., a number of other countries, in a pretty unrelated fashion. And yet these paths come together in this room today, and each of us pretty confidently expected the others to be here.

From the moment I leave my house in the morning to go to work, if not before, my life is at the mercy of my ability to figure out what people are going to do. If I'm walking or on my bike, scores of people will drive by who are capable of killing me with their cars. And yet, when I come to corners, a glance at the driver slowing down or speeding up, making eye contact or looking the other way, is usually enough for me to decide whether to cross or stay put. I put my life on the line like this almost every day, and I've made it to my sixties. Quite amazing, really. I must possess a pretty good way of telling what's going on with people, and predicting what they will do.

We are able to deal with people effectively because we are all in possession of one of the great intellectual accomplishments of humans, which I'll call "the person theory". The person theory isn't a scientific theory, but rather, a large, often vague, picture of how humans work that relate inner states and aspects of brains and central nervous systems—called "the mind" in the person theory—to one another and to observable stimuli and behavior. This theory allows us to make imprecise but quite helpful predictions about what people will do and gives us guidance as to how we ought to treat them. This theory has many virtues. But it has its limits. The person theory is based on a system of *indirect classification*. We describe the internal states of humans and predict their actions based on the functions those internal states perform when a well-functioning person is connected in the normal way with the external world. The theory breaks down, both descriptively and normatively when these conditions are not met. The Case Studies put

some of these limits on display.

The key features of this theory are *intentionality*, *local rationality*, *autonomy*, *identity* and *self*. We rely on the theory to construct identities, both of ourselves and others. Our understanding of ourselves and others appeals to these identities. When they break down our understanding of persons, both of ourselves and others, breaks down. When this happens, we have to take a step backwards, and try to understand what is going on in new ways. I'll first describe what I take to be the key concepts of the Person Theory, and then try to step back, and understand the Case Studies, where in various ways the person theory breaks down.

## 2 The Basics of the Person Theory

### Intentionality

“Intentionality” is a philosopher’s term for a variety of mental states and activities that we are able to describe in terms of facts about and possibilities for the rest of the world. I’m talking about some of the most ordinary and ubiquitous concepts that we use to describe people, such as belief, desire, knowledge, hope, and fear. Thinking in terms of these concepts is second nature to us, but when looked at closely, they are rather puzzling and amazing.

I assume you believe that Sacramento is the capital of California. Because you have this belief, you will act in certain ways. If asked what the capital of California is, you will move your lips and vocal chords in such a way as to emit the sound, “Sacramento”. The belief must be something inside of you, capable of being a partial cause of neurons firing and muscles moving.

But look at how we describe or identify this state. We don’t talk about neurons or synapses. We talk about Sacramento and California. We describe your belief by what the world has to be like for it to be true. Your belief is true *if* Sacramento is the Capital of California, false if some other city is. So think about it. There is something in your head, a belief. This belief has a property, being true. It has that property because of a fact about two things, Sacramento and California that are thousands of miles away. Isn’t this amazing? How can this be?

The missing elements, that are implicit in the way we talk about beliefs, are your *ideas*. They are in your head. At least, according to the person theory, they are in your mind, and according to modern science, the mind is the brain, or at the very least, the states of mind are determined by the states of the brain.<sup>1</sup> You have an idea *of* Sacramento, and an idea *of* California, and an idea of a city's being the capital of a state. So the picture begins to make sense. If these ideas are related in your mind in one way, you believe Sacramento is the capital of California. If they are related in another way, you believe it is not. Your belief is true if the relation among your ideas *corresponds* to the relation between Sacramento and California.

But this just raises the question, what is involved in an idea, like my idea of Sacramento, being *of* something? What is the relation between the aspect of my brain we call my idea of Sacramento, and Sacramento, the city thousands of miles away?

I got my idea of Sacramento when I was a kid, years before I saw the city. The connection between my idea and Sacramento went through the words in my geography textbook, the words my teacher used, the connections her words had to other people and texts and their use, and eventually to people that had some pretty direct relation with Sacramento, like seeing it. The idea was connected with certain abilities and bits of information. I learned to find Sacramento on a map. My new Sacramento idea became properly associated with idea of being the capital of (I already knew that Lincoln, where I lived, was the capital of Nebraska), and California. These connections were very complicated. As complicated as these connections are, we find it useful to talk about the brain and its ideas in terms of the things to which they have these complicated relations.

I'm connected in much the same way to all sorts of things, including, say, Aristotle and Plato, who have been dead for a couple of thousand years. It's really pretty miraculous that I can think about Aristotle, and have the sort of thought I express with "I wonder if Aristotle really liked Plato all that much, and if they ever hung out together and just talked about women or sports, or if they just had a student teacher relation that was sort of formal."

But for our purposes the main thing to note is that these complicated connections terminate in ideas in my brain that are in some sense components of my thought, something local, in my head, and capable of combining with other states to move my muscles and produce action.

### **Local Rationality**

---

<sup>1</sup> I argue for the former view, the identity theory, the mind is the brain and mental states are brain states, in *Knowledge, Possibility and Consciousness*.

The person theory doesn't just involve describing people in this sort of indirect way, based on their complicated connections to the world. We think that we can predict what people will do, and explain what we ourselves do, in terms of these intentional states. Let's return to the driver. I decide at a glance that he sees me and can tell that I intend to cross the street. I assume that he has some combination of the following desires: not to run over people, not to disobey the law, not to get a ticket or get sued. His desires, like his beliefs, I describe not in terms of neurons and synapses and the like, but in terms of things that might or might not happen in the future. And then I make a big assumption. I assume that because of having the belief that I am a pedestrian and want to cross, and that there is a stop sign there, and that I have the right of way, and that it would be against the law to interfere with my pass across the street, much less to run over and maim or kill me, and having the desires to not hurt me, to obey the law, to not get a ticket, and to not get sued, he will remain stopped. I infer what he will do, what muscles will contract, what limbs will move, what pedals will get pushed, and what, as a result of all that, he will cause the car to do. I assume that what he does will make sense, given his beliefs and desires. I assume that we will act in such a way so that his desires will be satisfied, if his beliefs are true. So I take my life in my hands, and cross the street in front of a huge SUV that could easily crush me.

Our expectations are based on taking people to be rational agents. The rationality is limited. We are not like Mr. Spock on Star Trek. Emotions get in the way of rationality. The processes by which we come to have beliefs, settle on goals, and make decisions may be very stupid and hardly deserving of the term "rational". A poll taken in Tennessee before the 2004 election showed that about 25% of the voters thought that Bush was in favor of taxing the rich heavily, and about 25% thought that Kerry favored federal support for parochial schools. These people were not very well-informed, but they were probably still *locally* rational. If they wanted to tax the rich, they probably voted for Bush; if they wanted federal aid for parochial schools, they probably voted for Kerry. Given the beliefs they had at the time, and that were operative in their deliberations and decisions, and their desires, urges, appetites, goals, and the like, they behaved in a way that would bring about what they want, or at least improve the odds of it happening. If one's beliefs are false, actions based on them will probably fail, even if one is locally rational. If one's beliefs are true, actions based on them will probably succeed, if one is locally rational. The beliefs, whether false or true, *close the gap* between what a locally rational person wants and what he or she does. *If* the beliefs are true, the action will promote their goals.

As I said, local rationality allows for a lot of stupidity. It also allows for bizarre and inexplicable desires. But however wild or irresponsible a person's desires and however misinformed their beliefs, we expect that what they do will be something that will advance at least some of their desires given their beliefs. This fundamental expectation forms that basis for our ability to interact with other humans. Without it, we would be lost.

## Respect for Autonomy

The person theory is not only a way of describing people's minds and predicting what they will do. It also has a moral side; it is a prescriptive theory as well as a descriptive one; it tells us how we should *treat* persons.

We believe that within certain broad limits, people should be allowed to act as their own beliefs and desires, their own reasons, dictate. If we want to change the way they act, we should change their reasons, and we should do so in ways that respect their intentionality and rationality, and fits at some level into their own projects and goals. Sometimes this is called "respecting autonomy" or "treating people as ends in themselves." This requires that if we want to influence what they are going to do, we should work to change their beliefs and desires in ways that respect their rationality, by providing them evidence that things may be different than they think. This isn't sufficient for respecting autonomy. Threats, for example, work by changing the structure of a person's desires, but often do not respect autonomy. The threat creates a local reason for them to do something that without the threat wouldn't fit into their own structure of goals and beliefs. I don't really want to hand the thief my wallet. But his threat to do my bodily harm creates a local reason for doing so.

Immanuel Kant said that moral people treat other people, at least some of the time, as ends not as means. This basically means that we treat them in ways that respects their own desires and goals as legitimate, and try to help them achieve these goals, even if doing so means putting our own goals on hold. Sometimes we treat people as means to achieving our own goals, but the moral person shouldn't do this all the time. Making people do something they don't want to do, even if we do so by exploiting their intentionality and local rationality, usually doesn't respect their *autonomy*, their right to act in accordance with their own vision of the good life.

We sometimes legitimately force or induce people by threats or bribes to do something that doesn't fit with their own goals. We do it "for their own good." This is justified if what we make them do or prevent them from doing really serves their own more important goals. Sometimes it is justified because it fits goals that they have had, or we are sure will have, even if they don't have them right at the moment. My child doesn't want to do his homework, doesn't see it of any value. I'm quite confident at some point in the future he will have goals that will benefit from his being well-educated, or at least not a complete ignoramus. This is legitimate. But what about a college student, a more-or-less adult? Based on years of experience and the incredible wisdom that comes with getting old, I may be certain that a student will be better off taking a couple of courses in computer science than simply devoting all of his units to the humanities. I may try to persuade him of this. I may be certain that if I convince him, he will come back later and thank me. Still, it's his decision. We ought to respect the autonomy of adults, even when they are autonomously doing something they will be sorry for.

## **Diminished and Fractured Selves**

The philosophers at this conference have been asked to comment on four cases that represent a partial breakdown of the person theory, and prevent those who care about the people involved from dealing with them as fully autonomous people. It is a partial breakdown, because the victims of Alzheimer's, frontotemporal dementia, severe apathy, and steroid psychosis strike us as persons, and continue in many ways to be treatable as such with a certain amount of success (in Alzheimer's this depends on how far along the disease has progressed). And yet something is wrong. Their rationality is diminished; the structure of their intentional states is skewed or incomplete; we are inclined to think we should violate their autonomy for their own sakes.

In each case some person acts in a way that deviates from the norm or paradigm of human behavior. They seem to fall short, in one way or another, of being autonomous persons. The philosopher's contribution to understanding such cases is not medical or neurological, but consists in trying to articulate the norms from which these cases deviate in a helpful and analytic way, to connect the situations with our ordinary ways of understanding and dealing with people, and to help trace the implications that this deviance should have, or should not have, in the way we deal with them. But before doing this, we need to say more about the person theory.

### **3 Personal identity and the Identities of persons**

#### **3.1 The Philosophical Problem of Personal Identity**

The traditional problem of personal identity for philosophers is this: under what conditions are Person A and Person B one and the same person? This can be a practical problem because we have inadequate knowledge of events. The practical problem of personal identity often arises in the judicial system. The prosecutor claims that the defendant, the person sitting in the courtroom, is the very same person who committed the crime, at a different place and a different time. The problem confronting the jurors is one of knowledge, of knowing the facts; it is, as philosophers say, epistemological.

If the jury had a complete video of everything that happened in all the relevant parts of the world — maybe this would require more than a video, perhaps some assemblage of hyperlinked digitized videos produced by a system video cams spread throughout America as a part of some future edition of the Patriot Act — they could probably be quite sure of the right answer. They would just rewind the video until they got to the crime, follow the movements of the criminal on the video or linked videos covering different the different regions of the world into which he wandered, and see if the criminal ended up coming into the courtroom and sitting at the defense table.

The philosopher is more likely drawn to what might be called metaphysical issues, issues that may remain after all of the facts are, in some sense, known.

Suppose that as the jury follows the career of the criminal, call him Roscoe, he does the following. He goes to a completely up to date brain science facility, where brain scientists have developed a technique for duplicating brains. The hope is that a person with some brain deterioration can have a new brain manufactured, made of sounder material, which will be psychologically indiscernible from the original. That is, when replaced, the new brain will give rise to the same beliefs and desires and memories and intentions as the old one; the headaches will disappear and the once inevitable strokes won't occur, but the intentionality will be the same as before. Roscoe has his brain duplicated. He has his original brain and his body destroyed and the duplicate brain put into a different body, Jeff's body. Jeff has just been declared brain dead, although his other organs are in fine shape. The criminal actually did this just to confuse things and make it hard to trace his movements. He swears the neurosurgeons to secrecy, but they don't cooperate.

The survivor of this operation leaves the hospital and ends up in the courtroom. He admits having memories, or at least something very much like memories, of committing the crime. But his lawyer claims that the criminal actually slipped up, and committed suicide. A human being is an animal, and this is a different animal, a different human. The defendant is actually Jeff, with a brain transplant. He is no more the criminal than he would be had he gotten the criminal's liver or heart. What we have here, the lawyer argues, is Jeff, a man who had a terrible injury, and who, though saved by a miracle, has lost all of his memories, in their place having delusions of a criminal past. Jeff is to be pitied, not punished. He calls some philosophers as expert witnesses (paying them less, no doubt, than other expert witnesses charge)—Bernard Williams say.<sup>2</sup>

The prosecutor is undeterred. He also calls expert witnesses, perhaps John Locke or Sydney Shoemaker. They explain that our concept of a person is not really a concept of an animal, but of a certain sort of informational action system, one that our person theory fits. These philosophers maintain that the person theory actually gives us a new concept of a continuing thing, one that conceivably could breach the bounds of bodily identity. Persons are systems that pick up information from experience, develop and sustain goals, and apply the information to achieve their goals. Such systems require a certain causal basis, some hardware on which the relevant data is stored and the relevant programs run. Usually this is provided by a single human body. But that is not a necessary requirement. Look, he may intone, we recognize the possibility of having the same person without having the same body when we talk of survival in heaven or hell, or reincarnation. These may be religious fantasies, but they show that it at least makes sense to have the same person when we don't have, in any ordinary sense, the same body or the same animal. Our criminal figured out a way of surviving the death of his body. The defendant is not Jeff, with a new brain and delusions, but Roscoe, with a new body and a duplicate brain.

---

<sup>2</sup> Essays by Williams, Locke, and Shoemaker on personal identity can be found in Perry, 1974.

Whether personal identity is just a species of animal identity, or something special based on the person theory, these philosophers focus on an important aspect of the theory that goes beyond intentionality and local rationality. We do not expect people, ourselves and others, to be merely locally rational, responding, in a given situation, to the impulses and desires that arise in them at that time, and the beliefs they can pick up by looking and listening to what is happening. Persons accumulate information about the world, and retain it, and they accumulate goals, aspirations and desires, and form long-term intentions and plans that structure their lives. They have enduring traits of character and personality. They have identities.

When a psychologist or an ordinary man (i.e., not a philosopher) talks about the identity of a person they do not have in mind mainly something that could be decided by fingerprints or a driver's license picture, but an enduring structure within the person, his or her own individual combination of beliefs, goals, habits, and traits of character and personality, the pattern that as we might say, *makes* the person who he is.

Of particular importance is the sense the person has of himself. What properties does this person think are true of him? Which ones are most important to him? How does he see this as fitting into a narrative of his life? A psychologist might have a person rank the properties he or she takes himself or herself to have in importance. Which properties can they not imagine not having? Can this man imagine being a woman? Would it matter a lot? Can this philosopher imagine being an accountant? Can this neuroscientist imagine being a philosopher? Does this mother find it incomprehensible that she should not be a mother, or is it an accident in her life? Would being different in these ways destroy a person's sense of who she or he is, and fracture the narrative of her or his life? Or could they be accommodated within the basic picture of himself that the person has? The most important, basic, inalienable facts about a person are more or less what the psychologist might think of as his or her identity.

When an Alzheimer's patient or a victim of frontotemporal dementia shows no interest in things that have long been important to her, such as cleanliness, caregivers may be distressed in large part because it is so out of character, because the person they once knew would be so shocked and ashamed at their current behavior. How do we respect such a person's autonomy? By limiting ourselves to the intentional structure that now motivates the person? Or by supplying for them what they cannot supply for themselves, commitment to their own past values? And often it is this threatened infidelity to our selves, rather than the pain and suffering connected with the later stages of the disease, that is likely to be most frightening to us who contemplate ending up as Alzheimer's patients, or to those who have been diagnosed with the disease but are in the early stages.

With the apathetic person, we may feel they have lost their identity, or lost touch with their identity, not in the strict and literal sense, for it is clearly the same person who was once energetic and is now apathetic, and they will take themselves to be that person. But the person's very own structure of values and beliefs and priorities seems no longer to be

able to motivate action. Who the person is, is no longer a guide to what the person will do. It is the idea that we may be helping the apathetic person return to their true identity, in the psychologist's sense, that might seem to justify so dramatic an intervention as is contemplated in the case study.

#### 4 Selves and the Sense of Identity

A word we often use in connection with a person's identity is "self". Some philosophers think of selves as rather mysterious immaterial entities. Sometimes selves are identified with the souls of Christian theology, or the essential natures that are passed along in reincarnation, or some noumenal object that exists beyond normal space and time, outside of the causal realm, and joins, in some Kantian way, with the primordial structure of reality to create the world as we know it. I don't think such mysterious notions of the self are required to understand the person theory. I think that a self is just a person, thought of under the relation of identity. But that sound mysterious enough, so let me explain.

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 "teaching" to some student. A father is just a person, thought of under the relation of *being the 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 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 basic concept of self is not of a special kind of object, but as a special kind of concept, that we each have of ourselves.

We each have a very special way of thinking about our self, that is, thinking about the person who we are, via the relation of identity. We have a *self-notion*, a concept of ourself as ourself. I want to say a bit about this key concept, about a person's sense of who they are, of their own identity, for this is a central part of the person theory.

Perhaps it's a little unclear what I'm looking for. Sometimes the best way to find something is to first consider a case where it is absent, and then see what is missing.

#### **Castañeda's war hero**

Now, a sort of paradigm case of someone who doesn't know who they are, and in that sense lacks a 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 maybe surgery, or maybe just time.

I'll use an example from the great late philosopher Hector-Neri Castañeda. He imagines a soldier — call him Bill — who having performed many brave deeds in a certain battle, is injured, loses his dog-tags, and awakens with amnesia. Not only does he not know who he is, no one else does either. He is clearly a soldier, 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 he is dead and his body unrecovered somewhere. He is awarded many medals posthumously.

For the time being let's concentrate on Bill, lying in the hospital, not knowing who he is. 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. Also, since he doesn't know his own name, he can't exploit *other people's* knowledge of who he is. He can't exploit public sources of information about himself. This is something we all rely on. If I forget my phone number, I can 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.

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. If 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.

Why do I say "normally"? Maybe 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 whom it is they are getting information about the immediate environment of visually. Most of us don't have to do that.

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" J. Carroll Nash 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 Nash 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 Nash'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 Nash 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."

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 that he is the war-hero, that his dissertation is actually, in a sense, 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.

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 Locke emphasized was 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 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 what one did and thought isn't *like* remembering what someone else thought and felt. And in the case of others, there are always the questions of *whom?* I remember someone eating poison oak, but was it Elwood? But if I remember eating poison oak, it was me that was doing the eating.

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-notion, his own conception of who he is. 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, but he won't remember doing them.

Also with respect to the future, I can know what you are doing, what you intend to do, 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 our own future..

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.

The notion that Bill was able to have of himself, even when he didn't know who he was, was his *self-notion*. Self-knowledge, in the ordinary sense, is knowledge of ourselves attached to our self-notion. Knowing facts about the person you happen to be, as Bill did when he wrote his dissertation, isn't enough.

If we know who we are, if we know our own names, we can 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 what we might call our *public identities*. This is the shared conception of us that others have. It is what our mothers and fathers and sons and daughters and colleagues and bosses and employees 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. For many issues, it is a better source of information about ourselves than any normally--self-informative method of knowing.

In fact, for many of us, perhaps 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 notion 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.

As we construct our public identities, we rely on the help of others. Public identities are a bit like works of art, or publications; they are accomplishments that take on a life of their own.<sup>3</sup> And of course they need not be unique. 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, and uncles and aunts and parents and each other; a somewhat different person in the eyes of my colleagues; and so forth. My self-conception, the picture of myself that animates me and explains how I act and react, may change subtly, or not so subtly, in different situations.

So I have a sense of my own identity. Here we see this other, psychological, use of “identity”. What is my identity? It is my own self-concept, the things I think hold true of me. A lot of this information I get from present perception: I think I am sitting in a chair, typing on a laptop, listening to Dixieland music, looking out the window at a rainy day. Some of it I have from memory. And some of it I have what others have told me about myself, and from applying general information about people to myself.

To reiterate:

- Each person has a special, dedicated, notion, his self-notion. This notion collects information acquired in normally self-informative ways, knowledge about his own mental and bodily states, and about the world around him is like, and what he has thought and done in the past, and will do, or at least plans to do, in the future.
- Our self-notions also serve to collect information we get about ourselves in other ways, as long as we recognize that it is ourselves that the information is about. I read in the email notice of the conference what time I will be giving a paper, and where. I pick up information about myself under the name “John Perry” which is the same way that others get information about me.
- Normally we expect a person to have a very complex self-concept, full of things that he has learned about himself in the past, both in normally self-informative ways and as a result of what others tell him about himself. We expect his desires and goals to be based not simply on urges and needs that he has now, that he can discover by present

---

<sup>3</sup> See Borges’ short story, “He and I”.

feeling and introspection, but also on memories of the past and goals adopted in the past.

#### 4.1 Two pictures of the self

Now I want to contrast two pictures of the self. One seems to be held by many philosophers and perhaps also by economists. I call this *the fully rational agent*. The other seems more accurate to me; I'll call this *the multiply centered self*.

Mr. Spock of *Star Trek* seems to have been a fully rational agent. When he 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 multiply centered self is based mainly on me, but I think it fits everyone I know somewhat better than the fully rational agent picture. I think it is also a more useful picture for considering diminished and fractured selves.

My goals and beliefs combine into clusters or centers, that vie for control of me. Which ones are operative, and among the ones that are operative, which ones are dominant, is a very situational thing. In part, I respond to what others expect. And my emotions play a large role, in pushing various centers forward, or making them recede, from consciousness and control.

Consciousness and control are not the same thing. For example, I'm at a department meeting. Someone says something which I interpret as a put down. I become angry. The anger "takes over". I respond in a way that is locally rational, given my strong desire not to be put down, not to let those who put me down escape unpunished, and to see them wince and squirm. But other centers, other clusters of desires and beliefs, look on appalled. Part of me, a little voice, urges restraint. "Calm down. You can be almost certain you won't lose anything by shutting up. You're not a child on a playground. You

are going to be sorry for this outburst. Etc. Etc.” This part of me continues thinking, even as I am uttering angry words. It has lost control of the agency, or at least that part of it that controls the mouth that is speaking. Another cognitive cluster is in charge. 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 dominant cluster. They are present; they are motivating the ineffective little voice. But they have nothing to do with what I am saying out loud. 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 and the like.

What we seem to have here are two different complexes of desires, beliefs, habits, etc., each associated with the same self-notion, producing different actions at the same time. One is producing verbal behavior; the other is producing thoughts and trying to take over the spoken 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 body that holds on to it, 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 desk full 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? Most of us, most of the time, keep some kind of order among our competing centers of agency.

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. Self-deception plays an important role here. I misremember my outburst, trying to interpret it as something that serves goals that actually had no role in motivating it. “I thought that these things needed to be said forcefully, in order to assure that other department members, less able to defend themselves, won’t get treated similarly in the future.” I say this first to my colleagues, and eventually believe it myself.

Emotions are feelings, and often feelings that overwhelm us, and divert us from single-minded rationality. Sometimes they get in the way; sometimes they save our lives. Emotions often serve to divert us from inappropriate deliberation, and move us more quickly to action. Fear says, “Don’t deliberate, flee.” Sometimes a good idea, sometimes not.

I don’t really think Mr. Spock would be all that good of a First Mate. What gets him out of bed in the morning? What gets me out of bed in the morning? Part of me says, “Get

up. Lots to do. Enjoy the sunrise. Get in an extra hour of work.” Part of me says, ”Who cares. Stay in bed.” What decides? Deliberation? No, just a slow shift, not under my rational control, between caring about getting something done more than staying warm and comfortable — or, sometimes, the other way round.

### **The three self-notions of Eve**

The competing centers model of the self postulates a single self-notion, with various centers competing for attention and control. Sloth, indifference, self-deception and occasional bursts of self-discipline allow their coexistence.

But although we naturally expect that there is one self-notion or self-concept to a customer, there is no logical guarantee of this. That is, we expect that there will be one notion where all of the information gained in normally self-informative ways will be found, and various parts of which will govern all actions that are normally self-effecting. But there is no contradiction in having more than one self-notion.

This seems the natural way to regard cases of multiple personality. Consider, e.g., the classic “Three faces of Eve”. Here we seem to have three self-notions, Eve White, Eve Black, and Jane. Each self-notion stores some of the normally self-informative knowledge, and each is partially responsible for self-effecting action. But the distribution of knowledge is rather complex:

- Eve White is in charge of action most of the time, and during those times gets knowledge about Eve gained through perception.
- Eve Black is in charge of action only occasionally — leading to episodes that surprise and confuse Eve White and lead her to seek therapy. She is also aware of “what happens to Eve White”. That is, she gets information about the thought and actions of both the Eve White and Eve Black personalities. But she keeps track of what Eve White sees and thinks and does in a way separate from how she stores knowledge about what Eve Black sees and thinks and does.
- Jane is aware of everything, but only very occasionally is in charge— only when the therapist makes a determined effort to get her to “come out”.

Here we have one human, but three ‘personalities’. Our ordinary expectations of how knowledge will be stored and action governed are upset. Eve’s self-conception had to vary, to make sense of what was expected of her in different situations, and how she had to think of herself to survive in these situations. The pressure was so great that she lost

the ability to sustain a single self-notion.

The person theory is a good theory, but one that is fragile. It's like an economic theory, that is based on well-working first world economies, but may fall short in explaining dysfunctional third-world economies. It's like our understanding of our computers, and how to operate them, that works fine when the disk works fine, the software is sound, and the electricity steady. For most of us, a good bit of the time, we can think of our computers in terms borrowed from the person theory. The computer knows how to do this and that; it wants to please me; if I ask it to find a file, or send some mail, or download a file from the web, it will happily do so. Most of us operate computers with only the vaguest sense of what goes on inside, and yet everything we do depends on affecting what goes on inside. The pull down menus with their English commands, the mouse, the keyboard, and all the wonderful *interfaces* computer engineers have designed allow us to control the internal states in terms of their functions and their connections with external things. If I click on the bookmark that is *of* The New York Times Crossword webpage, my computer will take me there. I rely on this *of-ness*, this intentional approach to my computer. When things go wrong I am lost. I turn to the computer expert. She fixes it, or gives me an estimate of what it will cost to do so. If the cost is too high, I throw it away. I have no interest in trying to operate the computer in any ways other than the intentional interface that I have learned permits.

But we can't throw people away. When they grow dysfunctional, and we can no longer use the person theory to understand them and influence them, we have to try other ways. The person theory breaks down; what do we put in its place?

## **5 The Case Studies**

### **5.1 Alzheimer Disease**

With Alzheimer's patients a number of questions naturally arise, from different perspectives, especially the patient himself, the caregivers, including family and close friends and the person who knows that he has Alzheimer disease, or anticipates getting it, and feels fear.

I should say that my father died of Alzheimer disease about six years ago, and I'm probably more influenced by my memories of that than by Mr. Jones case study. They are not dissimilar, except my father was a good deal less pleasant at the various stages of the disease than Mr. Jones' is described as being. I've thought a lot about Alzheimer's disease since that time. What I would have to say about frontotemporal dementia would be pretty similar, so I won't separately discuss that case.

There are different stages of Alzheimer's. In the early stages, when memory and functional ability decline, but Mr. Jones is still able to recognize family and deal with life more or less adequately, even though he is not the steady, reliable professional type that he seems to have been most of his life, there seems little need to bring in the heavy machinery of personal identity theory or self-theory to describe him. He is a person, the same person who was once a Fortune 500 Executive. He has a self, probably still more integrated than many, with goals and memories stretching back for years. It is a diminished but not a fractured self.

In the later stages of Alzheimer disease, one may be tempted to say more radical things. Perhaps Mr. Jones, when he no longer seems to remember much of anything, is not really the same person who contracted the disease? One might use a Lockean memory theory to bolster this conclusion. Perhaps he is not really a person any more, not coming close enough to the ideal of being a rational agent?

I don't think either of these things is the right thing to say. The kind of memories Mr. Jones is missing are the ones that connect with verbal abilities; using names, recalling facts, and the like. He still reacts to his wife in a way he doesn't react to others. Even if this were not so, even if he didn't discriminate between family, short term paid caregivers, and strangers, I don't think any reasonable theory of personal identity would say we have a new person here. Since the progress of the disease is gradual, so is the loss of memory. At any point, until the last stages, there will be memory links, however attenuated and partial, that will take us, step by step, to the person's past.

There is a stage in any wasting disease that affects the brain where one is inclined to say that even if there is some brain function left, we no longer have a person; we no longer have the person we knew and loved, but only the same live human being. But I think in the case of Alzheimer's we have some temptation to say this early on. When the patient ceases to recognize us, or even before, when they cease to treat us the way they always have, cease to be able to play the role in our lives that we are used to them playing, we suffer loss. And we must also often make difficult decisions that seem to violate autonomy: taking away driving privileges, or even walking privileges; not letting them write checks; supervising phone conversations which are necessary because of the various predators that prey on old people by phone and mail (and now email); ignoring temper tantrums. Our loss may be easier to deal with if we imagine that the person we knew has ceased to be; the violations of autonomy may be easier to stomach if we think we no longer have a functioning person.

But I don't think, until the very last stages of the disease, that these descriptions are accurate. We are dealing with persons who have changed dramatically, and who, as the disease progresses, have diminished selves, and whose autonomy can be legitimately violated for their own sakes. But we are not dealing with non-persons or different persons, in the strong sense. Their identities change, in the psychological sense, but they are the same person, in the philosophical sense, however diminished and changed.

There is one aspect of things that may be especially troubling for the caregivers. I'm not sure what to say about it, but I'll try to say something.

Suppose that Mr. Jones, like some Alzheimer patients, comes to prefer a daily routine which would have appalled him at earlier stages of his life. He doesn't like to shower, or get dressed, and doesn't much care if he smells bad. (This is what happens to Mr. Smith, the frontotemporal dementia patient.) Imagine this at a stage when he is still continent and still remembers his children and grandchildren, somewhat vaguely. But when they come to visit he has no inclination to get cleaned up. He dresses so casually that the younger grandchildren are shocked.

Mrs. Jones is absolutely and correctly certain that this behavior would appall the old Mr. Jones. Indeed, in the early stages of the disease, it was one of the things that he feared most. Perhaps he even made her promise to force him to shower and look decent, even if she had to hire a strong-arm nurse and he protested vigorously.

Now she doesn't know what to do. The suffering both physical and emotional that Mr. Jones will endure if she does what is necessary to get him presentable is real, and hard for her to take. The thought of her husband appearing so disheveled is hard for her to take. She thinks she is obliged to honor her promise to the earlier Mr. Jones. But sometimes she thinks he had no right to ask of her, that she should coerce another human being to do what he so much hates to do, especially when that human being has such a limited realm of autonomy, having been stripped, by intervention or incapacity, of much of what he enjoyed in life.

What should Mrs. Jones do? Well, it probably depends on the specifics of the situation, maybe on the specifics of each visit; which grandchildren will be coming, and exactly how bad Mr. Jones smells that day. But in general, there are a couple of things to remember.

When we do things for children, "for their own sakes," what we have in mind is a later stage of the same child, whose purposes will be served by having had their autonomy limited at the earlier stage. Billy may not want to practice the piano, but won't Bill the teenager be happy that he can understand and play music? There is danger in such thinking. Bill the teenager may wish he had been left alone. But the basic reasoning seems acceptable.

But there is an asymmetry; between treating the present stage of a person in some way they don't want to be treated, doing it for the sake of a later stage, and doing it for the sake of an earlier stage. Mrs. Jones needs to keep this in mind. In fact Mr. Jones never will look back and thank her for getting him cleaned up. If she does it, she is not doing it for Mr. Jones sake, but for her own sake and for the sake of her children and

grandchildren.

But this isn't the whole truth. She may be doing it for the sake of his memory, or for the sake of his public identity. She is sad that her own memories of her husband as a good husband and father and successful businessman have been blurred if not obliterated by the later stages of his life. She may wish that her children's memories not be similarly affected, to the extent that she can help it, and that her grandchildren not have this helpless smelly old man as their main memories of their grandfather. How does one balance these feelings with the reluctance to make a living, feeling person do things they don't want to do? Things which are of no risk to them?

Suppose Mr. Jones had been a successful academic, and had written a very fine book that had gone through many editions and was considered a classic in its field. At some medium stage of the disease he completes a revision, which really converts the book into a hopeless mess, something that would humiliate the earlier Mr. Jones. He desperately wants it published. Mrs. Jones, we can suppose, has the powers of attorney to prevent this. Should she allow this? Does she have a right not to allow it?

Here the unalloyed person theory says: well, it's *his* book. He has the right to do with it as *he* pleases. But this doesn't seem right. An analogy might be the case of moral responsibility. In some sense I am morally responsible for all the wrongs I have committed. But I don't feel guilty about things I did as a kid, or even a young man, to the extent that they happened long ago, and that the values, desires, weaknesses, and the like that led me to do them no longer operate within me. I may continue to feel responsible for making amends if damage continues.

It seems that in the same way that I have a right to release myself from guilt (which is a bit different than responsibility) from things done long ago, due to another identity, I cease to have rights to positive things I have done. Do old artists have a right to destroy the brilliant work they did when they were young? Do old rich men have the right to destroy the trusts and bequests they made when they had a sounder identity?

In a sense, a person's public identity, the sense the world has of her, is an accomplishment. A life well-lived, or well-lived enough to instill pride in and provide a model for younger members of one's family, and to produce, on balance, pleasant memories in family friends and associates, is a considerable accomplishment. And we in fact care greatly about such things. It is a commonplace that Presidents care about the "verdict of history." This can seem a little odd. The verdict of history means, when you come right down to it, what predicates will be attached to your name in history books after you die. For most of us there will be no history books, but there will be memories of those who survive you, stories your children tell to your grandchildren, maybe a little money that makes someone's life easier. We care about such things. Do our later selves, our diminished and fractured selves, have a right to destroy them, simply in virtue of

being later stages of the same person? I think Mrs. Jones has good reasons to not let this happen. She shouldn't suppose that she is doing it for Mr. Jones' sake, in the way one guards a youth from indiscretion for the sake of the adult to follow. But she also shouldn't suppose that the diminished Mr. Jones who is present, has unbridled rights to destroy the public identity, the memories and opinions of him that others have, that his earlier self worked so hard to build up.

How does one rationally think of the possibility, or probability of suffering from Alzheimer's disease? Dying of Alzheimer's is not pretty or pleasant, but I don't think it is the dying that is likely to provoke the most fear. It is the idea of being so diminished, so unlike the way we want to be, of being such a burden on loved ones, of doing embarrassing things, that horrifies us. And yet, when we are at that stage, we may have enjoyable lives. My father, in a home in Harlingen, Texas, spent a year where he mainly thought of himself being in Italy, where he served in World War II, in some sort of rest home, with attractive and attentive Italian nurses. It was not an unpleasant life, but it was one he would have looked forward to with horror, had he known of it in advance.

I think it is rational to fear the inconvenience and hardship one will cause one's loved ones, the drain one may put on their energies and bank accounts, the embarrassment one may cause. And one can reasonably fear the damage that anyone will do to one's public identity, the memories people have of one, and the record of accomplishment for which one has striven. The fact that the person doing the damage may be oneself doesn't diminish, but augments, the fear. Still, these things can be handled, with proper planning, if one is fortunate enough to know what is coming and has the resources to make plans.

Some people fear dying. Hume suggested that they think of all the years before they were born, and so didn't exist. That wasn't so bad. Why should non-existing after death be something to fear, if thousands of years of nonexistence prior to birth have left no bad memories? Perhaps one should compare one's future as an Alzheimer's patient similarly. It wasn't so bad being a child, with the diminished self and lack of autonomy that small children have.

## **5.2 Deep Brain Stimulation**

In 2006 Charles Garrison has Deep Brain Stimulation to treat his apathy, a syndrome associated with his Parkinson's disease. A fairly large number of electrodes are placed in his brain. The result is a radical change, but not a change back into the sort of person he was before the onset of severe apathy. The once successful, diligent, conscientious, shy, family-oriented quiet Mr. Garrison became extremely outgoing, gregarious, desirous of a great deal of attention, not very diligent at work, and only superficially interested in his

family. He also became a Democrat and an ardent environmentalist.

What happened to Mr. Garrison? One thing is clear, the second is not so clear.

Hume said that reason is, and ought to be, a slave to the passions. What exactly he meant I don't know, but I like the remark in almost any possible interpretation. The relevant one is that deliberation alone won't get us out of bed. As I lie there in bed I may know how important the day's activities are. Perhaps I can convince myself that if I make it into work and teach my Introductory Philosophy Class, young minds will be influenced that will some day save the world from tyranny and bigotry. That won't, by itself, get me out of bed. Some unfathomable something has to happen. I have to, at that moment, care about that goal. It is not enough to realize that teaching the class is what I ought to do, what I want to do, what the world wants me to do, and expects me to do. Some passion, some caring, has to kick in. Dwelling on my goals, my duties, all the good that can come from getting out of bed may give rise to such passion. Usually it does. But it doesn't constitute it. To any set of desires and beliefs that present themselves to one as one's own cherished hopes and goals, one can always consistently add, "so what," turn over, and go back to sleep.

When one is depressed the "so what" addendum crops up all the time. One is tempted to think that some further beliefs or goals have been added to one's usual repertoire, and changed what it is rational for one to do. All is vanity. The world is going to end before long anyway. No one really cares what I do. I can't really make a difference. But that doesn't seem to be phenomenologically accurate to me. The depression comes first; the rationales for not caring come along later. The basic fact is, one just doesn't care, or doesn't care enough, to get out of bed, to grade the papers, start the coffee, or whatever. The cure, for the normal range of cases, is not more deliberation, but alarm clocks placed in the next room, automatic coffee pots, wives that give you a kick, cats that sit on your head until you get up and feed them. And antidepressants help. Without passion, or at least irritation, reason is impotent.

What seems clear is that the deep brain stimulation did for the more severely apathetic Mr. Garrison what my cat does for me. He wasn't depressed, according to his psychiatrist. It's hard for me, in my own mind, to imagine what it is like to be apathetic without being depressed. Of course, there is being lazy, another thing I have personal knowledge of. At any rate, the therapy got him moving, or, given that it hasn't happened yet, will have gotten him moving.

The means are a bit frightening, because one imagines how a technology that involves putting electrodes deep in one's brain could be exploited and abused. Slippery slopes seem all around us. If electrodes to cure apathy now, why not to cure liberalism or an inordinate love for philosophy in the near future? But the cure for slippery slopes is carefully chosen constraints: good shoes, sand, ropes, barriers, and things like that. The

cure is not forgoing the good things that lie along the slope, if we don't slip.

The deep brain stimulation is frightening because it is so clearly an intervention that doesn't involve any appeal to reason, to Mr. Garrison's autonomy. But he seems to have agreed to it. There doesn't seem to be much to object to.

But was it a success? That brings up issues that are not so clear. Mr. Garrison is no longer apathetic. He is full of energy. But is he the same Mr. Garrison? Well, in the strict and literal sense of personal identity, he surely is. Mr. Garrison changed, but he continued to exist. But where did these new needs, values, habits, commitments, and concerns come from? We wanted a quiet, diligent, Republican, family man, and got a talkative, self-absorbed, environmentalist Democrat.

What we would like in this case is knowledge of what happened to Mr. Garrison that tied in with the person theory. Within that theory we seem to be able to frame to alternative interpretations of what happened, and we'd like to know which one was correct.

One interpretation is this. Mr. Garrison, like all of us, was a disorderly complex of intentional states, held in some sort of equilibrium by external expectations and internal negotiations. The values exhibited in his environmentalism, democratic party membership, story-telling, superficial interest in his family, loquaciousness, and the like were always in there, losing the battle for dominance. This is not to say that his diligent, Republican, family-man life was a facade or a charade. That too was part of who he was, a part whose dominance depended on a variety of things that, for one reason or another, have ceased to hold sway. The new Mr. Garrison is one who might have emerged, under different circumstances of bodily health, for quite different reasons: using alcohol, ceasing to use alcohol, undergoing psychotherapy, finding Jesus, who knows what. The Parkinson's disease, the apathy, and the therapy have shaken up the equilibrium that shaped the Mr. Garrison his family knew and wanted back. But the new Mr. Garrison is authentic, a new equilibrium among competing centers. With a little work, maybe his family can get some of the old values and cares to reemerge.

The other interpretation is that the therapy did more than reinvigorate Mr. Garrison's passion for living. It destroyed values, commitments, and traits that he had, and replaced them with new ones, with no previous basis in his self. It would not necessarily follow that the therapy was mistake. One might encourage an apathetic kid to join the Navy. One expects, indeed hopes, that his values and beliefs and approach to life will be shaken to their roots and changed. One may hope that what emerges from this experience will still like to hunt and fish and go to Giants games with his father and vote Democratic, but maybe not. Who knows what will happen. A vegan, Republican, Dodgers fan will still be a big win.

It may be that asking which sort of case we have with Mr. Garrison is a mistake. Maybe we are pressing questions that make sense within the person theory to a situation to which it simply doesn't apply. It is not a case beyond the limits of human understanding. But perhaps it's not a case within the limits of the human understanding that the person theory can provide, either descriptively or morally. I wish I knew.

### 5.3 Steroid Psychosis Case

There are two things about the John Fast case that I want to comment on. First the decisions he makes, while not under the influence of steroids, to begin and then to resume taking them. Then there are his actions while under the influence of steroids, or while suffering from steroid psychosis.

Why would someone who was as successful as John Fast begin taking steroids to boost his strength, and why would he do it surreptitiously? Well, this isn't so mysterious really. Like most American boys, I wanted to be an athlete badly. I went through the various attractive options, baseball, basketball, football, learning at different ages the futility of hoping for even modest success. I played football for a small college team my freshman year---small enough that I had thought---wrongly---that I would have a chance of getting some playing time. After the last game, the coach put his arm around me and said, ``Perry, you are small, but you are slow. Why don't you go try being a philosophy major, or something for which you are remotely suited." Good advice.

From the point of view of a totally frustrated athlete, John Fast's baseline life seems like one that one wouldn't want to mess with. But with a little imagination, I can relate to it. Suppose there were some pills that made one just 10% mentally quicker, 10% better able to understand the complex arguments and subtle insights and occasional logical formulae that my philosophical colleagues throw my way. Suppose there were even some dangers. I might become 10% less charming and affable, 10% less supportive of my children. It sounds like a pretty good deal, really. I might find it very attractive, if I really believed it would work. Not that I think I have an incredible surplus of charm and affability. But I must admit I find myself in situations in which I think I am at the limits of my intellect, or somewhat beyond, more often than I find myself at the limits of my charm and affability. With 10% more brainpower, and 10% less charm and affability, I think I would be more like my philosophical heroes than I am now. Come to think of it, are you guys working on any pills like this?

But where were the desires that motivated John Fast to this momentous decision, all the years when he was becoming a successful athlete, model team member, and admirable husband and father? Deep in his psyche, waiting for the right situation to provoke them, I suppose.

Why was he private about it? Because he was to some extent ashamed of his desire, and his willingness to use steroids. We aren't told very much about the factors that would determine whether his initial decision was rational. Steroids aren't, or didn't used to be, illegal, at least some of them. I gather they are often effective. Some say that they are responsible for the great home-run hitters of the recent era, except for Barry Bonds, of my beloved Giants, whose late-career strength is completely due, I'm sure, to clean living and weight training. As a casual reader of sports-pages, I wasn't aware that steroid psychosis was a risk. Maybe John Fast wasn't either---the first time, although he certainly was the second time he decided to use steroids.

But his use was surreptitious. This suggests that either he didn't think his reasons would bear the scrutiny of his wife and coaches, or, probably more likely, that he didn't want to explain his reasons. John Fast perhaps had some dissatisfaction with the nature of his success, perhaps a dissatisfaction that only emerged with awareness of the possibility of taking steroids and becoming stronger. We all have desires that we don't want to disclose to others, or even fully admit to ourselves; at least I do, and I suppose others do too.

So his initial decision to take steroids falls more or less within the sorts of things we can understand with the person theory. It is motivated by beliefs that probably had some connection with reality and information available to him, and desires that were not that dissimilar to those that had motivated him throughout his life.

His later decision is harder to understand. Did he really believe that the effect of steroid use would be different his time? Or was there something about the period when he was on steroids that we haven't grasped, that made him long to return to that condition? That would be intelligible to us with mood enhancing drugs, drugs that produce an experiential high. The steroids seem to have produced in John Fast a self-image high. Although he was in fact not successful on the field, and was making his family miserable, it seems that he was pretty happy. But later, when he is, it seems, convinced that his good feelings about himself and his performance, and his attribution of jealousy to others, was a delusion, why would he want to return to his delusional state?

This question is similar to one that arises with sort of mirror image cases, people who are more successful and even thrive when they stick to their drugs for bipolar disorder or paranoid schizophrenia, but refuse to take the drugs. In the past, at least, this was intelligible because the drugs in question have unpleasant side effects. And people with bipolar disorder, I gather, often miss the manic phase.

With John Fast the question that arises in my mind is whether both the initial and subsequent decision to use steroids was based on some deep dissatisfaction with the nature of his life and his success, a deep-seated desire to be a superstar, even if it meant the loss of the ties to team and family that, one would have thought, would be an incredible source of satisfaction, pride and comfort. Was this desire so deep that it was preferable to return to the state where he seemed to himself to be such a superstar, even knowing that it was a delusion? Or so deep that during the period between steroid use,

that he didn't really believe what he was told about his life as a steroid user? Or was it rather that the first period of steroid use changed him in ways that make it hopeless to try to understand his subsequent decision in terms of anything like the preference structure that he had before?

What he tells us, in Ending One, is that after resuming steroid use John Fast believed he was a better athlete and a better person, and liked his "new self" better than his old self. Were the good will of his teammates and family never so important to him as it seemed? Was he always a would-be prima donna, trapped in the life of a good family man and team-player, with no way to get out until the steroids came along?

The person theory, as I described it, is powerful but limited. It involves a very indirect way of getting at the electrical and chemical phenomena that move us. It provides no way that I know of for getting at answers to the questions of the sort I have raised, and perhaps there are no answers.

What of John Fast while on steroids? For those of us who are limited in our interactions with other human beings to the power that the person theory gives us---who have no recourse to medical knowledge, no ability to write prescriptions---there is a paradox in dealing with people who are addicted to various things. I've had more experience than I would have chosen to have dealing with people addicted to alcohol and to cocaine. The paradox is that while one is obviously dealing with someone to whom the person theory does not apply, as it should, the best thing to do seems to treat him or her as autonomous humans who must take complete responsibility for their actions---to avoid becoming a "co-dependant".

We naturally say, in such a case, that the person isn't who they used to be. But, for reasons like those given above with the Alzheimer patient, I see no reason to deny personal identity. The problem isn't that we have a new person, but that the old person has chosen to get themselves into a situation where the character traits one once could rely on--- honesty, a certain amount of prudence, feelings for those they love --- will play any significant role in their decisions. But in dealing with them one must assume that memories, cares, affections, and traits of old still have some presence in their psyche. One simply has to do one's best to make sure that when the immediate effect of the cocaine has abated, the person will then remember the negative effects of the cocaine use, not softened by efforts of co-dependants, with such horror as to produce a new equilibrium among the completing motivating complexes. Sometimes it works, and sometimes it doesn't, as the two ending to John Fast's story suggest.

## 6 Appendix. Identity: Basic Concepts and Distinctions

### 6.1 Identity versus Similarity

The concept of personal identity is a special case of what is sometimes called numerical identity. The relevant concept of identity is expressed in various ways, “are identical,” “are one and the same” etc. If X and Y are identical, in this sense, there is just one thing that is both X and Y. So if the cows Bossie and Trixie are one and the same, if they are identical, then there is just one cow, called both “Bossie” and “Trixie.” English is confusing in various ways. Almost all the words for numerical identity are also used to convey similarity. E.g., imagine now we have two cows, one named “Bossie” and the other named “Trixie”. They are both Guernseys, both give the same amount of milk, both are somewhat ornery when milked. We might say, “Bossie and Trixie are the same,” meaning that they are very similar or very much alike. Maybe the farmer liked Bossie so much he went looking for as similar a cow as he could get, he wanted one just like Bossie. We might say he wanted the “same cow” or even “the identical cow.”

Note that in the numerical sense of identity, the sense in which there is just one thing, the idea of identical twins makes no sense. If they are identical, they are not twins; if they are twins, they are not identical. “Identical” in “identical twins” doesn’t mean numerical identity, but similarity, or perhaps coming from a single egg.

#### Logical Properties of identity

From now on I’ll use “identity” in the sense of numerical identity unless I indicate otherwise. The logical properties of identity are simply consequences of the idea of just being one thing. For example, if you just have one thing, it has all the properties it has:

- If  $x$  is identical with  $y$ , and  $y$  has property  $P$ , then  $x$  has property  $P$ . [The indiscernibility of the identical]
- Further: If  $x$  is identical with  $y$ ,  $y$  is identical with  $x$  [symmetry]
- If  $x$  is identical with  $y$ , and  $y$  is identical with  $z$ , then  $x$  is identical with  $z$  [transitivity]
- Everything is identical with itself, that is, for all  $x$ ,  $x$  is identical with  $x$  [Reflexivity]

## Identity and time

The Greek philosopher Heraclitus got tenure for saying that you can't step in the same river twice, because new waters are always flowing in. This is deep and profound, but not quite right. Of course you can step in the same river twice, although as you do so, you won't be stepping in exactly the same water, at least if the river is flowing at any rate at all.

If we just say that when you step in the same river at two different times, it will not be exactly similar as it was before, it doesn't sound quite so profound.

Suppose that the Cayster is full of muddy water on Monday, but clear on Tuesday. Then don't we have the problem? How can one river have different properties at different times, given the principle we called the indiscernibility of the identical?

We just have to be careful. The same river has the property of containing muddy water Monday, and also the property of containing clear water Tuesday. If we include the time in the property, there is no problem.

Even if we speak in the normal tensed way, there is no problem if we are careful. The principle of the indiscernibility of identicals implies:

If  $x$  and  $y$  are identical,  $x$  has all the properties  $y$  has, and  $x$  had all the properties  $y$  had, and  $x$  will have all the properties  $y$  will have.

But it doesn't imply:

If  $x$  and  $y$  are identical,  $x$  *has* all the properties  $y$  *had*...

Suppose Heraclitus stands in the clear Cayster on Tuesday, and says, "I stepped in this very river, the identical river, one and the same river, yesterday, and then it was muddy." From this he can infer that the river he is standing has clear water, and had muddy water, the day before, and that the river he stood in yesterday had muddy water in it then, and has clear water in it now." But he shouldn't have concluded that it can't be the same river he is standing in today as he was standing in yesterday.

## Continuity, Causation and Identity

The concept of identity is applied to everything, concrete objects, abstract objects (like numbers and properties), contrived objects (like the sequence consisting of the Eiffel Tower and Bob Dylan), clouds, wind currents, and so forth.

Persons belong to the very general category of concrete things, things which have a position in space and endure through time. It is often thought that the identity conditions of concrete things amounts to spatial temporal continuity. Why is the coin in my pocket now the same one I put in there this morning? Because there is a spatiotemporal continuous path that stretches from spatiotemporal position of the coin this morning to the spatiotemporal position of the coin in my pocket now, and every point along this path is or was occupied by a coin. This is certainly something we at least expect of concrete objects, and it is the reason we usually think we can establish identity by establishing such a continuous history —as we imagined our jury doing in the case of Roscoe the criminal.

For most concrete things there is also an element of direct causality built into our concept. Technology provides a lot of ways of giving the illusion of a concrete thing although what we really have is a spatiotemporal connected succession of different things, made to provide the illusion of a single thing. For example, if I type an “s” in this file, and then go back and insert from some spaces, I will think of the “s”s I type as moving to the right along the line. This “s” isn’t really a single concrete thing, but a succession of things made to give the appearance of a single thing. (Of course, it is a single *succession*, but a succession isn’t a concrete thing, and a succession of “s”s isn’t an “s”). The similarity of the first s and the second s doesn’t result from the usual sort of direct causality that makes a concrete thing look pretty much the same from instant to instant, even if it moves a little. Rather, one thing is annihilated and another put in its place by the editing program. I’ll call this virtual identity.

In the case of the succession of letters, we don’t really have continuity. That would require that between any pair of “s”s in the series there was another overlapping “s”. So maybe we can distinguish between virtual identity and real identity on that basis. On the other hand, are we sure that we really have continuity in the case of ordinary objects? It isn’t really something we can observe. If the scientists at SLAC or CERN tell me that we don’t really have temporal continuity, but that the careers of physical objects turn out to be as full of little temporal gaps, I’d have to believe what they say. So I think we need to appeal to a concept of direct causality. The position, and the characteristics, of each successive stage of a physical object are explained by the position and characteristics of the earlier stage.

Ordinarily, we expect concrete things to change in gradual ways, unless there is a particular event that results in a lot of changes. I expect the coin in my pocket now to look pretty much the same as the one I put in my pocket this morning. Of course, if some time during the day I took it out and put it on a railway track and let a train flatten it, then it won’t. That change will be explained, however, by the way the coin was, and the pressures that the train exerted on it. The careers of concrete objects have a characteristic shape, each stage explained by how they were, and what happens to them.

This applies to humans in their physical aspects. You will expect me to look pretty much

the same tomorrow as I do today, unless I get run over by a car or undergo cosmetic surgery or something like that. The similarity isn't due to some outside agency or program that is keeping track of how the successive John Perry's the worlds sees ought to look. It's just a consequence of the way people develop. Of course if people look too much the same as earlier stages of themselves, where the earlier stages are considerably earlier, that also requires explanation. If the person in question lives in Los Angeles, we assume cosmetic surgery.

Our concept of the identity of a person fits into this general scheme, even though the psychological characteristics of persons, their beliefs, desires, and traits, are much different sorts of properties than the shapes and sizes and appearances of (merely) physical things. Even is we adopt a Lockean theory of personal identity, and allow that we may have the same person even if we do not have the same animal, or as Locke puts it, allow that we can have the same person when we don't have the same man, we will have not abandoned entirely our ordinary conception of identity as grounded in the direct causation of basic similarities or explicable differences in the important properties of the object in question.

## References

- Castañeda, Hector-Neri , “ ‘He’: A Study in the Logic of Self-Consciousness,” *Ratio* 8 (1966): 130–157.
- [] Perry, John. 2000. *Knowledge, Possibility, and Consciousness*. Cambridge: MIT Press.
- [] Perry, John (ed). 1974. *Personal Identity*. Berkeley: University of California Press.
- [] Perry, John. 2002. *Identity, Personal Identity, and the Self*. Cambridge: Hackett Publications.
- Thigpen, Corbett H. and Hervey M. Cleckley.& Cleckley, The 3 Faces of Eve. Popular Library (1957). Revised edition, Acata Graphics (1992).