JOHN PERRY

Personal Identity, Memory and the Self

In this paper, I will contrast two concepts of personal identity, the philosophical and the psychological. Then I'll develop my account of self-knowledge. In the course of this I'll explain the concept of memory that I think is crucial to developing our sense of ourselves as persisting beings, and contrast it with other ways we can remember things about ourselves. I'm very interested in the light our colleagues here in the Center involved in the memory project can shed on this distinction.

The Philosophical problem of personal identity

Professor John Perry,
Department of Philosophy,
University of Stanford, USA.
CAS Fellow 2003/04.

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 of 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 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, if 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.[1]

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.

Identity

Let's spend a little time on the concept of identity.

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 (transivity)
- 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 amount 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 spatio-temporal continuous path that stretches from spatiotemporal position of the coin this morning to the spatio-temporal 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 spatio-temporal 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 some spaces, I will think of the "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 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 pres-

sures 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 if 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.

Psychological identity

Now I want to consider a different, and perhaps more common, sense of "personal identity."

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 persons'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.

Selves and the sense of identity

A word we often use in connection with a person's identity is "self". The concept of self involves both philosophical and psychological identity.

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 sounds 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* 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.

Perhaps its 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.

Castaneda'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 Castaneda. 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 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.

Types of 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 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 is always the question of *who*? 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 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, 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.

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. 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 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 lap-top, 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 from what others have told me about myself, and from applying general information about people to myself.

Let me close by reiterating the basics of my account of self-knowledge:

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

- All of our actions are ultimately motivated by information that is stored in, or connected with, our self-notions. This information can motivate normally self-effecting actions. And all of our actions, however unselfish, and however remote we intend their consequences to be, come down to moving our limbs and other bodily parts in various ways, intended to bring about wider and wider changes, in virtue of the circumstances we are in.

References

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.

Notes

1 Essays by Williams, Locke, and Shoemaker on personal identity can be found in Perry, 1974.