Personal Identity over Time*

Theodore Sider

November 21, 2004

1 The concept of personal identity

On trial for murder, you decide to represent yourself. You are not the murderer, you say; the murderer was a *different person* from you. The judge asks for your evidence. Do you have photographs of a mustachioed intruder? Don't your fingerprints match those on the murder weapon? Can you show that the murderer is left-handed? No, you say. Your defense is very different. Here are your closing arguments:

> I concede that the murderer is a righty, like me, has the same fingerprints as I do, is clean-shaven like me. He even looks exactly like me in the surveillance camera photographs introduced by the defense. No, I have no twin. In fact, I admit that I remember committing the murder! But the murderer is not the same person as me, for I have changed. That person's favorite rock band was Led Zeppelin; I now prefer Todd Rundgren. That person had an appendix, but I do not; mine was removed last week. That person was 25 years old; I am 30. I am not the same person as that murderer of five years ago. Therefore you cannot punish me, for no one is guilty of a crime committed by *someone else*.

Obviously, no court of law would buy this argument. And yet, what is wrong with it? When someone changes, whether physically or psychologically, isn't it true that he's "not the same person"?

*This is the first chapter of *Riddles of Existence: A Guided Tour of Metaphysics* — a metaphysics book for students and non-philosophers, written by Earl Conee and me, forthcoming with Oxford University Press.

Yes, but the phrase "the same person" is ambiguous. There are two ways we can talk about one person's being the same as another. When a person has a religious conversion or shaves his head, he is *dissimilar* to how he was before. He does not remain **qualitatively** the same person, let us say. So in one sense he is not "the same person." But in another sense he *is* the same person: no other person has taken his place. This second kind of sameness is called **numerical** sameness, since it is the sort of sameness expressed by the equals sign in mathematical statements like "2+2=4": the expressions '2+2' and '4' stand for one and the same number. You are numerically the same person you were when you were a baby, although you are qualitatively very different. The closing arguments in the trial confuse the two kinds of sameness. You have indeed changed since the commission of the crime: you are qualitatively not the same. But you are numerically the same person as the murderer; no *other* person murdered the victim. It is true that "no one can be punished for crimes committed by someone else." But "someone else" here means someone numerically distinct from you.

The concept of numerical sameness is important in human affairs. It affects whom we can punish, for it is unjust to punish anyone numerically distinct from the wrongdoer. It also plays a crucial role in emotions such as anticipation, regret, and remorse. You can't feel the same sort of regret or remorse for the mistakes of others that you can feel for your own mistakes. You can't anticipate the pleasures to be experienced by someone else, no matter how qualitatively similar to you that other person may be. The question of what makes persons numerically the same over time is known to philosophers as the question of **personal identity**.

The question of personal identity may be dramatized by an example. Imagine that you are very curious about what the future will be like. One day you catch God in a particularly good mood; she promises to bring you back to life five hundred years after your death, so that you can experience the future. At first you are understandably excited, but then you begin to wonder. How will God insure that it is *you* in the future? Five hundred years from now you will have died and your body will have rotted away. The matter now making you up will, by then, be scattered across the surface of the earth. God could easily create a new person out of new matter who resembles you, but that's no comfort. You want *yourself* to exist in the future; someone merely like you just won't cut it.

This example makes the problem of personal identity particularly vivid, but notice that the same issues are raised by ordinary change over time. Looking back at baby pictures, you say "that was me." But why? What makes that baby the same person as you, despite all the changes you have

undergone in the intervening years?

(Philosophers also reflect on the identity over time of objects other than persons; they reflect on what makes an electron, tree, bicycle, or nation the same at one time as another. These objects raise many of the same questions that persons do, and some new ones as well. But persons are particularly fascinating. For one thing, only personal identity connects with emotions such as regret and anticipation. For another, *we* are persons. It is only natural that we take particular interest in ourselves.)

So how could God make it be *you* in the future? As noted, it is not enough to reconstitute, out of new matter, a person physically similar to you. That would be mere qualitative similarity. Would it help to use the same matter? God could gather all the protons, neutrons, and electrons that now constitute your body but will then be spread over the earth's surface, and form them into a person. For good measure, she could even make this new person look like you. But it wouldn't *be* you. It would be a new person made out of your old matter. If you don't agree, then consider this. Never mind the future; for all you know, the matter that now makes up your body once made up the body of another person thousands of years ago. It is incredibly unlikely, but nevertheless possible, that all the matter from some ancient Greek statesman has recycled through the biosphere and found its way into you. Clearly, that would not make you numerically identical to that statesman. You should not be punished for his crimes; you could not regret his misdeeds. Sameness of matter is not sufficient for personal identity.

Nor is it necessary. At least, *exact* sameness of matter isn't necessary for personal identity. People survive gradual changes in their matter all the time. They ingest and excrete, cut their hair and shed bits of skin, and sometimes have new skin or other matter grafted or implanted onto their bodies. In fact, normal processes of ingestion and excretion recycle nearly all of your matter every few years. Yet you're still you. Personal identity isn't especially tied to sameness of matter. So what *is* it tied to?

2 The soul

Some philosophers and religious thinkers answer: the *soul*. A person's soul is her psychological essence, a nonphysical entity in which thoughts and feelings take place. The soul continues unscathed through all manner of physical change to the body, and can even survive the body's total destruction. Your soul is what makes you *you*. The baby in the pictures is *you* because the very same soul that now inhabits your body then inhabited that baby's body. So

God can bring you back to life in the future by making a new body and inserting your soul into it.

Souls might seem to provide quick answers to many philosophical perplexities about identity over time,[1] but there is no good reason to believe that they exist. Philosophers used to argue that souls must be posited in order to explain the existence of thoughts and feelings, since thoughts and feelings don't seem to be part of the physical body. But this argument is undermined by contemporary science. Human beings have long known that one part of the body — the brain — is especially connected to mentality. Even before contemporary neuroscience, head injuries were known to cause psychological damage. We now know how particular bits of the brain are connected with particular psychological effects. Although we are far from being able to completely correlate psychological states with brain states, we have made sufficient progress that the existence of such a correlation is a reasonable hypothesis. It is sensible to conclude that mentality itself resides in the brain, and that the soul does not exist. It's not that brain science *disproves* the soul; souls *could* exist even though brains and psychological states are perfectly correlated. But if the physical brain explains mentality on its own, there is no need to postulate souls in addition.

Also, soul theorists have a hard time explaining how souls manage to think. *Brain* theorists have the beginnings of an explanation: the brain contains billions of neurons, whose incredibly complex interactions produce thought. No one knows exactly how this works, but neuroscientists have at least made a good start. The soul theorist has nothing comparable to say, for most soul theorists think that the soul has no smaller parts. Souls are not made up of billions of little bitty soul-particles. (If they were, they would no longer provide quick answers to philosophical perplexities about identity over time. Soul theorists would face the same difficult philosophical questions the rest of us face. For instance: what makes a soul the same over time, despite changes to its soul particles?) But if souls have no little bitty soul-particles, they have nothing like neurons to help them do their stuff. How, then, do they do it?

3 Spatiotemporal continuity and the case of the prince and the cobbler

Setting aside souls, let's turn to scientific theories, which base personal identity on natural phenomena. One such theory uses the concept of **spatiotem-**

[1] Perhaps too quick: could we not raise the question of identity over time for souls?

poral continuity. Consider the identity over time of an inanimate object such as a baseball. A pitcher holds a baseball and starts his windup; moments later, a baseball is in the catcher's mitt. Are the baseballs the same? How will we decide? It is easiest if we have kept our eyes on the ball. A **continuous series** — a series of locations in space and time containing a baseball, the first in the pitcher's hand, later locations in the intervening places and times, and the final one in the catcher's mitt — convinces us that the catcher's baseball is the same as the pitcher's. If we observe no such continuous series, we may suspect that the baseballs are different. Now, we don't usually need this method to identify a person over time, since most people look very different from one another, but it could come in handy when dealing with identical twins. Want to know whether it is Billy Bob or Bobby Bill in the jail cell? First compile information from surveillance tape or informants. Then, using this information, trace a continuous series from the person in the jail backward in time, and see which twin it leads to.

Everyone agrees that spatiotemporal continuity is a good practical guide to personal identity. But as philosophers we want more. We want to discover the *essence* of personal identity; we want to know *what it is* to have personal identity, not merely how to tell when personal identity is present. If you want to know whether a man is a bachelor, checking to see whether his apartment is messy is a decent practical guide; if you want to tell whether a metal is gold, visual inspection and weighing on a scale will yield the right answer nine times out of ten. But having a messy apartment is not the *essence* of being a bachelor, for *some* bachelors are neat. Weighing a certain amount and appearing a certain way are not the essence of being gold, for it is possible for a metal to appear to be gold (in all superficial respects) but nevertheless not really *be* gold. The true essence of being a bachelor is being an unmarried male; the true essence of being gold is having atomic number 79. For in no possible circumstance whatsoever is something a bachelor without being an unmarried man, and in no possible circumstance is something gold without having atomic number 79. All we require of practical guides for detecting bachelors or gold is that they work most of the time, but philosophical accounts of essence must work in all possible circumstances. The **spatiotemporal continuity theory** says that spatiotemporal continuity is indeed the essence of personal identity, not just that it is a good practical guide. Personal identity just *is* spatiotemporal continuity.

The theory must be refined a bit if it is really to work in every possible circumstance. Suppose you are captured, put into a pot, and melted into soup. Although we can trace a continuous series from you to the soup,

the soup is not you. After being melted, you no longer exist; the matter that once composed you now composes something else. So we had better refine the spatiotemporal continuity theory to read as follows: persons are numerically identical if and only if they are spatiotemporally continuous via a series of *persons*. You are connected to the soup by a continuous series all right, but the later members of the series are portions of soup, not people.

Further refinements are possible (including saying that any change of matter in a continuous series must occur gradually, or saying that earlier members of such a series *cause* later members.) But let's instead press on to a very interesting example introduced by the seventeenth-century British philosopher John Locke. A certain prince wonders what it would be like to live as a lowly cobbler. A cobbler reciprocally dreams of life as a prince. One day, they get their chance: *the entire psychologies of the prince and the cobbler are swapped*. The body of the cobbler comes to have all the memories, knowledge, and character traits of the prince, whose psychology has in turn departed for the cobbler's body. Locke himself spoke of souls: the souls of the prince and the cobbler are swapped. But let's change his story: suppose the swap occurs because the brains of the prince and the cobbler are altered, without any transfer of soul or matter, by an evil scientist. Although this is far-fetched, it is far from inconceivable. Science tells us that mental states depend on the arrangement of the brain's neurons. That arrangement could in principle be altered to become exactly like the arrangement of another brain.

After the swap, the person in the cobbler's body will remember having been a prince, and will remember the desire to try out life as a cobbler. He will say to himself: "Finally, I have my chance!" He regards himself as being the prince, not the cobbler. And the person in the prince's body regards himself as being the cobbler, not the prince. Are they right?

The spatiotemporal continuity theory says that they are *not* right. Spatiotemporally continuous paths stick with *bodies*; they lead from the original prince to the person in the prince's body, and from the original cobbler to the person in the cobbler's body. So if the spatiotemporal continuity theory is correct, then the person in the cobbler's body is really the cobbler, not the prince, and the person in the prince's body is really the prince, not the cobbler.

Locke takes a different view; he agrees with the prince and the cobbler. If he is right, then this thought experiment refutes the spatiotemporal continuity theory. Here is a powerful argument on Locke's side. Suppose the prince had previously committed a horrible crime, knew that the mind-swap would occur, and hoped to use it to escape prosecution. After the swap, the

crime is discovered, and the guards come to take the guilty one away. They know nothing of the swap, and so they haul off to jail the person in the prince's body, ignoring his protestations of innocence. The person in the cobbler's body (who considers himself the prince) remembers committing the crime and gloats over his narrow escape. This is a miscarriage of justice! The gloating person in the cobbler's body ought to be punished. If so, then the person in the cobbler's body *is* the prince, not the cobbler, for a person ought to be punished only for what he himself did.

4 Psychological continuity and the problem of duplication

Locke took the example of the prince and the cobbler to show that personal identity follows a different kind of continuity, **psychological continuity**. According to the new theory that Locke proposed, **the psychological continuity theory**, a past person is numerically identical to the future person, if any, who has that past person's memories, character traits, and so on — whether or not the future and past persons are spatiotemporally continuous with each other. Locke's theory says that the gloating person in the cobbler's body is indeed the prince and is therefore guilty of the prince's crimes, since he is psychologically continuous with the prince. As we saw, this seems to be the correct verdict. But Locke faces the following fascinating challenge, presented by the twentieth-century British philosopher Bernard Williams.

Our evil scientist is at it again, and causes Charles, a person today, to have the psychology of Guy Fawkes, a man hung in 1606 for trying to blow up the English Parliament. Of course, it might be difficult to tell whether Charles is faking, but if he really does have Fawkes's psychology, then, Locke says, Charles *is* Guy Fawkes. So far, so good.

But now our scientist perversely causes this transformation *also* to happen to another person, Robert. Coming to have Fawkes's psychology is just an alteration to the brain; if it can happen to Charles, then it can happen to Robert as well. Locke's theory is now in trouble. Both Charles and Robert are psychologically continuous with Fawkes. If personal identity is psychological continuity, then both Charles and Robert would be identical to Fawkes. But that makes no sense, since it would imply that Charles and Robert are identical to each other! For if we know that

$$x = 4 \quad \text{and} \quad y = 4$$

then we can conclude that

$x = y$.

In just the same way, if we know that

Charles = Fawkes and Robert = Fawkes

then we can conclude that

Charles = Robert.

But it is absurd to claim that Charles = Robert. Though they are now qualitatively similar (each has Fawkes's memories and character traits), they are numerically two different people. This is the **duplication problem** for Locke's theory: what happens when psychological continuity is duplicated? (Or triplicated, or quadruplicated. . .)

Williams chose spatiotemporal over psychological continuity because of the duplication problem. Before we follow him, let's think a little harder about spatiotemporal continuity. Just as a tree can survive the loss of a branch, a person can survive the loss of certain parts, even very large parts. You are still the same person if your legs or arms are amputated. Yet losing a part causes a certain amount of spatiotemporal discontinuity, since the region of space occupied by the person abruptly changes shape. Thus, "spatiotemporal continuity" should be understood as meaning *sufficient* spatiotemporal continuity, in order to allow for change in parts while remaining the same thing or person.

How much continuity is "sufficient" spatiotemporal continuity? Imagine that you have incurable cancer in the right half of your body but are healthy in the left. This cancer extends to your brain: the right hemisphere is cancerous while the left hemisphere is healthy. Fortunately, futuristic scientists can separate your body in two. They can even divide the brain's hemispheres and discard the cancerous half. You are given a prosthetic right arm and right leg, an artificial right half of your heart, and so on. You need no prosthetic right brain hemisphere, though, because the remaining healthy left hemisphere eventually functions exactly as your whole brain used to function. (Though fictional, this is not wholly far-fetched: the hemispheres of the human brain really can function independently when disconnected, and duplicate some — though not all — functions of each other.) Surely the person after the operation is the same as the person before: this operation is a way to save someone's life! But the operation results in a fairly severe spatiotemporal discontinuity, since the continuity between the person before and the person after is only the size of half the body. Moral: even the

continuity of only half the body had better count as sufficient for personal identity.

But now the spatiotemporal continuity theory faces its own duplication problem. Let us alter the story of the previous paragraph so that the cancer is only in your brain, but is present in both hemispheres. Radiation treatment is the only cure, but it has a mere ten percent chance of success. These odds are not good. Fortunately, they can be improved. Before the radiation treatment, the doctors divide your body — including the hemispheres — in two. Each half-body gets artificially completed as before; then the radiation treatment of the cancerous brain-halves begins. This gives you two ten-percent chances of success rather than one. But now comes the twist in the story: suppose the unlikely outcome is that *each* hemisphere gets cured by the treatment. So the operation results in two persons, each with one of your original hemispheres. Note that each is "sufficiently" spatiotemporally continuous with you, since we agreed that a half-person's worth of continuity counts as sufficient. The spatiotemporal continuity theory then implies that you are identical to each of these two new persons, and we again have the absurd consequence that these two new persons are identical to each other.

Each of our theories, Locke's psychological continuity theory and the spatiotemporal continuity theory, faces the duplication problem. A single *original person* can be *continuous*, whether psychologically or spatiotemporally, with two *successor persons*. Each theory says that personal identity is continuity of some kind. So the original person is identical to each successor person, which then implies the absurdity that the successor persons are identical to each other. How should we solve this problem?

Some will be tempted to give up on scientific theories and instead appeal to souls. Continuity, whether psychological or spatiotemporal, does not determine what happens to a soul. When a body is duplicated, the soul in the original body might be inherited by one of the successor bodies, or by the other, or perhaps by neither, but not by both. While this is a tidy solution, it is unsupported by the evidence: there still is no reason to believe that souls exist. It would be better to somehow revise the scientific theories to take the duplication problem into account. (If we succeed, we will still need to decide between psychological and spatiotemporal continuity, or some combination of the two. But set this aside for the remainder of the chapter.)

As we originally stated the scientific theories, they said that personal identity is continuity. We could restate them to say instead that personal identity is **nonbranching** continuity. Continuity does not *normally* branch: usually only one person at a time is continuous with a given earlier person. In such cases there is personal identity. But the duplication examples involve

branching, that is, two persons at a time who are both continuous with a single earlier person. So according to the restated theory, there is no personal identity in such cases. Neither Charles nor Robert is identical to Guy Fawkes. You do not survive the double-transplant operation.

Unlike the claim that the successor persons are identical to each other, this is not absurd. But it is pretty hard to accept. Imagine that, before the operation, you receive some good news: the left-hemisphere person will survive the division operation. Excellent. But now, if the modified spatiotemporal continuity theory is correct, then if the right-hemisphere person survives in addition, you will not survive. So it is *worse* for you if the right-hemisphere person survives. You must hope and pray that the right-hemisphere person will die. How strange! The news that the left-hemisphere person would survive was good; news that the right-hemisphere person would also survive just seems like more good news. How could an additional piece of good news make things much, much *worse*?

5 Radical solutions to the problem of duplication

Duplication is a really knotty problem! Perhaps it is time to investigate some radical solutions. Here are two.

Derek Parfit, the contemporary British philosopher, challenges a fundamental assumption about personal identity that we have been making, the assumption that personal identity is *important*. Earlier in this chapter we assumed that personal identity connects with anticipation, regret, and punishment. This is part of the importance of personal identity. The last paragraph of the previous section assumed another part: that it is very bad for you if no one in the future is identical to you. That is, it is very bad to stop existing. Parfit challenges this assumption that identity is important. What is really important, Parfit says, is psychological continuity. In most ordinary cases, psychological continuity and personal identity go hand in hand. That is because, according to Parfit, personal identity is nonbranching continuity, and continuity rarely branches. But in the duplication case it does branch. In that case, then, you cease to exist. But *in this case*, Parfit says, ceasing to exist is not bad. For even though you yourself will not continue to exist, you will still have all that matters: you will have psychological continuity (a double helping, in fact!)

Parfit's views are interesting and challenging. But can we really believe that utterly ceasing to exist is sometimes insignificant? That would require a radical revision of our ordinary beliefs. Are there other options?

We could instead reconsider one of our other assumptions about personal identity. The duplication argument assumes that if personal identity holds between the original person and each successor person, we get the absurd result that the successor persons are the same person as each other. But this absurd result follows only if personal identity is numerical identity, the same notion that the equals sign ("=") expresses in mathematics. We made this assumption at the outset, but perhaps it is a mistake. Perhaps "personal identity" is *never* really numerical identity. Perhaps all change *really does* result in a numerically distinct person. If so, then we would not need to say that branching destroys personal identity. For we could go back to saying that personal "identity" is continuity (whether psychological or spatiotemporal — that remains to be decided). In branching cases, a single person can stand in the relationship of "personal identity" to two distinct persons; that is not absurd if personal identity is not numerical identity. We would still need to distinguish mere qualitative similarity ("he's not the same person he was before going to college") from a stricter notion of personal "identity" that connects with punishment, anticipation, and regret. But even this stricter notion would be looser than numerical identity.

Can we really believe that our baby pictures are of people numerically distinct from us? That too would require radical belief revision. But sometimes, philosophy calls for just that.

Further Readings

John Perry's anthology *Personal Identity* is an excellent source for more readings on personal identity. It contains a selection from John Locke defending the psychological continuity view, a paper by Derek Parfit arguing that personal identity is not as significant as we normally take it to be, a paper by Thomas Nagel on brain bisection, and many other interesting papers. Perry's introduction to the anthology is also excellent.

> John Perry, editor, *Personal Identity*, University of California Press, 1975.

Another good book, also called *Personal Identity*, is co-authored by Sydney Shoemaker and Richard Swinburne. The first half, written by Swinburne, defends the soul theory of personal identity, and is especially accessible. The second half, written by Shoemaker, defends the psychological continuity view.

> Sydney Shoemaker and Richard Swinburne, *Personal Identity*, Blackwell, 1984.

Bernard Williams introduces the problem of duplication in this article.

> Bernard Williams, "Personal Identity and Individuation", in his book *Problems of the Self*, Cambridge University Press, 1973.