

Philosophical Conversations

An Interactive Newsletter

Vol. 5, No. 1

Spring, 2002

Department of Philosophy

Illinois State University

Philosophical Conversations is designed to stimulate thought and discussion, and keep you philosophically active. The format will be the presentation of a brief position paper to which responses are encouraged. In the subsequent issues selected responses may be published in addition to a new position paper. We invite you to respond to this position paper, either by contacting the author or the Department. (Addresses, mail and e-mail, and phone numbers and fax numbers are provided at the end of this issue).

A Puzzle about Punishment

by
Mark Siderits

There was recently a news story about a 75-year-old Chicago man who has been in prison since he was 17 on a murder conviction, and is now seeking clemency. This raises an interesting philosophical question: to what extent do we deserve to be punished for crimes we committed during a much earlier stage of our lives? We might think the answer is obvious: if someone is guilty of a crime, it makes no difference how long ago it was; since it's still just as much them, they still deserve the same punishment. But a certain philosophical theory of personal identity—what is called Reductionism—would challenge the premiss that it is necessarily just as much them. A Reductionist might say that sometimes what we think of as 'being the same person' can come in diminished degrees. If this is right, this might have an effect on what we say about cases like that of the convicted murderer.

A Reductionist theory of personal identity claims that the continued existence of a person does not involve the continued existence of some one thing, the self. Instead Reductionists say the continued existence of a person just consists in the occurrence of a series of causally connected psychophysical elements. Buddhists are Reductionists, but so are a number of modern philosophers like Locke, Parfit and Shoemaker. What would it mean to believe that I have no self? Some people think it would mean being like certain of neurologist Oliver Sacks' patients, namely those with an extreme form of Tourette's Syndrome. Such people lack any notion that their actions have consequences for them—that their past choices affect their present situation, and their present acts will have ramifications for their future. Instead they exist wholly in the present and give no thought to what might come tomorrow. Indeed it seems inconceivable to them that they will have a tomorrow. Now the reason some people think this is what it would be like to believe oneself to be without a self is that over the course of a lifetime all the observable parts of our bodies and minds go out of existence and get replaced. So if I thought there wasn't a self holding all these parts together, then I would think that the person tomorrow who is thinking different thoughts and has a few less hairs would be somebody else, not me.

The Tourette's Syndrome patients Sacks describes are pathological. So if this were what it is like to believe that there is no self, then anyone who believed that would be in a pathological state. But Reductionists usually claim that we would be better off if we believed what they take to be the Reductionist truth about ourselves. So something must be amiss here. To see what this might be, I suggest we consider an interesting mode of travel, teletransportation.

Suppose you want to go someplace exotic for a vacation. But now instead of visiting the Grand Canyon, you have the option of exploring some canyons on the newly opened resort on Mars. Until now the only way to get to Mars was by a 6-month trip on a spaceship—with all the discomfort of weightlessness and the hazards of solar radiation, collisions with rocks, etc. But now there's this newly proven technology that can beam you up to Mars in no time at all. What it does is scan your body and compile information on the complete state of every molecule. Then it sends that information to the receiving station on Mars, where a new brain and body that perfectly replicate the original are assembled out of new matter. Now the scanner here on Earth destroys the original. But the person who walks out of the receiving station on Mars looks just like you, has all your endearing personality traits, remembers all of your experiences, knows everything you know, etc. So as far as that person is concerned, the teletransporter was just a quick and painless way to get from Earth to Mars. Now suppose you know that the Teletransporter is reliable—indeed that it has a much better safety record than the spaceship company (the old Alaska Airlines). Which way would you choose to go: by Teletransporter, or by spaceship?

Those of you who remember the original Star Trek series will remember that the *Enterprise's* chief engineer Scotty wouldn't use the Teletransporter. He was afraid that when you were beamed down to the surface from the *Enterprise*, it left out something crucial. This in spite of the fact that Scotty knew that the resulting person always walked and talked and behaved and felt exactly like the original. So what is it that might have been left out? It's true that the heart, lungs, brain etc., of the person on Mars are not the original body parts but new ones that are just like the old ones. Is it that I have to have the same original heart in order to survive? Well, in that case it would never make sense to have a heart transplant when I have a life-threatening coronary condition. And similarly for all the other parts of our bodies: we would rather have a properly functioning duplicate than keep the original, if keeping the original meant death or even serious disability. So what Scotty is afraid gets left out by the Teletransporter is something it seems we could never miss. As long as there is a causal process that perfectly duplicates my body and my mental states, I will continue to exist even when all the original parts are destroyed. It would be irrational to fear Teletransportation and take the spaceship instead. The person who leaves the receiving station on Mars will be me.

To say there is no self is to say that there is no part of a person that must endure for that person to continue to exist. At any one time a person is made up of many different parts—all the cells that make up their body, and all the thoughts, feelings, memories and the like that make up their mind at that time. What the possibility of Teletransportation tells us is that all of these might go out of existence and yet the person continue to exist—provided something happens that causes the right kinds of replacements to come into existence. This is what ordinarily happens when cells of our body die—new cells are caused to grow and take their place. And we are beginning to understand how our brain states cause a similar kind of continuity in our mind.

The continued existence of a person thus involves a great many causal connections between the very many parts making up a person at one time and that person at another time. This raises a new possibility. Suppose that there is a malfunction in the Teletransporter while the information about your body and brain is being sent to Mars. The malfunction is like what happens when your car radio picks up two stations at once, so that your news program is punctuated by bits of Polynesian rap music. What the receiver on Mars gets is some of the information about your bodily and mental states combined with some information about the states of someone else. First suppose that there's only a tiny bit of overlap: the information is 99% about your body and mind, and only 1% about this other person. Thus the person leaving the receiving station on Mars will look 99% like you, will have 99% of your memories and personality traits, etc. In this case that person is so much like you that we surely ought to say that they are you. For cosmetic surgery might make me look very different, yet that would still be me. And there are many personality traits I used to have that I no longer do, yet it is still me for all that. A mere 1% change is not enough to make that person on Mars someone else.

Of course if it were the opposite—99% someone else's bodily and mental states, only 1% yours—we would say that you did not survive the malfunction. 1% of the normal causal connections is simply not enough for personal identity. Moreover, adding just a few more causal connections isn't going to turn a failure into a success. So if the resulting person on Mars were 98% like the other person and 2% like you, we would still say that you did not survive. But this process can continue: we can add a few more connections, and agree that those few aren't enough to make the all-important difference. The problem is that eventually we will get to a case where the resulting person is 70%, or 90%, or 99%, or even 100% a replica of you. If we continue to say in each succeeding case that the resulting person is not you, then we end up contradicting ourselves: surely the person who duplicates your bodily and mental states with 100% fidelity is you. So where did the resulting person stop being someone else and start being you? The difficulty is that we cannot say—any point we chose would be entirely arbitrary, and it seems wrong that something like whether or not you survive should turn on an arbitrary decision like this.

The same thing is true, though, of heaps. Suppose that we have a heap of 100 stones on the terrace, and we want the heap moved out under the locust tree. If 99 of them made it out to form a pile under the tree, and one more got added to that pile from somewhere else, we'd count that as a success. If, on the other hand, only one made it to the pile, which was then made up of 99 stones from somewhere else, and the other 99 from the terrace got scattered around, we'd say the heap under the tree wasn't the same heap as the one that had been on the terrace. But suppose the proportions were 70-30, or 60-40, or 50-50; then we might not know what to say. In this case, however, the fact that we don't know what to say in the middle cases is not so surprising. After all, it's not as if there really are such things as heaps in the world, over and above the stones, or grains of sand, or whatever it is that makes up the heap. The word 'heap' is just a convenient way to talk about a large number of things when they are all piled together. It's sometimes useful to have a way to refer to all the things piled together, and it would simply take too long to list each and every one. But it would defeat the purpose behind introducing the word 'heap' if we had an exact definition of how many stones it took to make a heap, or what proportion had to still be there for the heap under the tree to be the same one as the heap that was on the terrace. Instead we use the word in a rough and ready way, and aren't bothered when we come across borderline cases. In such cases we agree that there may not really be a fact of the matter as to whether or not something is a heap, or is the same heap. This is something we might not say about the individual stones: we think there has to be a definite fact of the matter as to whether this stone is the same one as the one I picked up earlier. But this is because we think stones really do exist, in a way that heaps do not. A heap is really just a useful fiction.

The Reductionist says that persons are likewise just useful fictions. At any one time, a person is made up of a great many bodily and mental states. Over time those states go out of existence and cause new bodily and mental states to arise to take their place. The word 'person' is therefore just a convenient way of referring to all the physical and mental states that make up this hugely complicated causal series. This is why it shouldn't surprise us when, in the case of the malfunctioning Teletransporter, there are borderline cases where there is no fact of the matter as to whether the resulting person on Mars is you or someone else. The bodily and mental states that make up what we think of as a person are really real. But when we think there also exists a person who has those states, this is a mistake. We only think this because we happen to have this handy way of referring to an entire series of such states, and we forget that the word is no more than a convenient device. The person who 'has' these body parts and these mind parts is no more real than the heap that 'has' those stones. There really are just the stones, arranged in a certain way. And likewise there really are just these bodily and mental states, going out of existence but causing new ones to come into existence, in a continuous complex causal series.

At this point you might be thinking that those people I talked about earlier were right to compare being without a self to the pathology of Tourette's Syndrome. If I believed there is no more to me than just a series of impermanent bodily and mental states, and there is nothing that holds them all together as 'mine', wouldn't I become completely dissociated from my past and future and end up trapped in the immediacy of the present? But to think this is to forget that the person was said to be a **convenient** fiction. This means that it can be useful for us to be able to collect together all the states making up a causal series and think of them as if they made up one thing. Think, for instance, about flossing your teeth, or getting a flu shot. Neither action is particularly pleasant, so there isn't any immediate reward for performing them. But if they don't get performed, eventually there will be a great deal of pain that could have been prevented. The best way to prevent that pain turns out to be getting the elements in each causal series to identify with the past and future elements in that series. This is what I do when I think that I should now floss my teeth in order to prevent the future pain of gum disease that will happen to **me** if I don't. Pain is a bad thing, and should be prevented. There will be less overall pain if each causal series learns to identify with the past and future stages of that series--learns to think of itself as a 'person'.

What Reductionists claim, in other words, is that while there really aren't any persons (any more than there really are any heaps), there are very useful results to be had if we think of ourselves and others as persons—as things that persist, that anticipate their future states, have feelings like remorse toward their past states, and carry desert for their actions. We can see how a system of punishing criminals fits in here. Crimes are actions that cause a great deal of overall pain, so the less crime there is the better. The criminal law in effect says that anyone who commits one of those acts deserves a certain punishment. Since we think of ourselves as persons, we identify with those future stages that

would suffer that punishment if we were to commit a crime and be caught, and this thought tends to deter us. Those who aren't deterred and are punished are less likely to repeat the act—because they identify with the future stages that would have to undergo the punishment again. So this is why we all think that criminals deserve to be punished: because we've all learned to think of ourselves and others as persons.

But as with most useful devices, it's always possible to improve the performance of this person idea with some fine-tuning. For it may be that we sometimes take this idea too seriously. Reductionists commonly see this happening in our attitudes toward our own death. They think the existential suffering that reflective people experience when they recognize their own mortality is the result of wrongly thinking that we have a self that is our essence. We can see how this might happen. Because it's so hard to socialize small children into thinking of themselves as persons, we tend to overdo it. The result is that we grow up with the belief that there is this little thing inside that is the true 'me' now, and it will continue to be 'me' in the future. And once we get used to thinking that way, it's terrifying to contemplate the possibility that there might come a day when it does not exist. Our lives would go better, Reductionists claim, if we could learn to take a little less seriously the idea that we are persons: continue to act like persisting beings with future interests (since this does make things go better), but stop believing in the fiction of an enduring self.

Is fine-tuning also possible with respect to our ideas about punishment? There are many different factors involved here, so to simplify things a bit I shall consider not that 75-year-old murderer, but another kind of case: someone serving a life sentence under a 'three strikes' law. Suppose the convict in question received this sentence at age 28 after a criminal career beginning, in adolescence, with a series of offenses that escalated in seriousness, and culminated in three felony convictions, the last being for armed robbery. Indeed suppose that in this case the law really did do what its sponsors claim such laws do in general: prevented a hardened criminal from inflicting harm on his victims through a life of crime. But now suppose as well that at age 75 this really is 'a completely different person'. Of course we shouldn't read too much into this phrase. The mind and body in that cell today are the outgrowth of the mind and body of the hardened criminal who was rightfully convicted 47 years ago. The prisoner remembers the experiences of the young criminal. He even acknowledges responsibility for his crimes; he feels real remorse for the suffering inflicted on his victims. But while this is true, still he can no longer identify with that young man. He can understand what led him to act as he did: what beliefs and desires were operating, what larger values and life-plans were at work. But those beliefs, desires, values and life-plans now seem utterly alien to him. His present mind-set is so at odds with that earlier one that he cannot now imaginatively reconstruct such things as the sadistic pleasure he once took in victimizing others. If he were told that he would once again become a person with just such a character (perhaps through treatment inflicted by some evil scientist), he would be utterly horrified; death would seem preferable to such a fate. In short, we might say, while they are the same person, the young criminal does not represent the real self of the old convict.

Reductionism tells us how to understand this way of talking. According to Reductionism, the continued existence of the same person involves a great many causal connections between the psychophysical elements existing at one time and the psychophysical elements existing at another time. As long as there is an unbroken chain of enough such connections, there continues to be the same person. But this allows for gradual change over time, so that in the end we may be confronted with something of a totally different character. We see this sort of process in biology all the time, for instance in the transformation of a tadpole into a frog. And it does occasionally happen with the character of persons as well. When it does, it might not be too misleading for the present person to refer to the earlier stage in their life both as 'me' and as 'not really my present self'.

Suppose that this is true of our 75-year-old convict. Now precisely because it may be to his advantage for us to believe this of him, we may be inclined to be suspicious here. So I will simply stipulate that this is true. Of course in the real world we seldom have the advantage of learning about people's characters by stipulation, so this may represent a problem in applying what we say about this case to real-world examples. But let's just see what would actually follow here. Does this man deserve to remain in prison until he dies, or may he be granted clemency and released now? Of course you can point out that he is the person who was given a life sentence, and deservedly so. But once we understand in a Reductionist way what his being the same person as the young criminal actually amounts to, do we still want to say that he is now just as deserving of the life sentence as he was then? Or should we say instead that since his present self is so very different from his earlier self, he is no longer as deserving of this degree of punishment?

MARK SIDERITS joined the Philosophy Department in 1980, after earning his B.A at the University of Hawaii and his Ph.D. at Yale, and has been the bane of ISU students' existence ever since. (It's possible that he figured in the childhood nightmares of some students even before he came here.) He teaches majors courses in the history of philosophy (Western and Asian) and metaphysics, as well as general education courses in applied ethics and feminism. Most of his research has been in Buddhist and Indian philosophy, and he's presented papers on those subjects to conferences on four continents. His second book, on reductionism about persons, will be out shortly. He's particularly notorious for regularly using (or at least mentioning) the f-word in many of his classes. Emily, his cat, loved him, but then she died. Since we should probably say something positive about him, we'll add that he does continue to ride his bike to school every day. Of course being a reductionist philosopher, he's always saying that wholes like bicycles don't actually exist, only the parts are really real. So whether that little bit of ecological virtue really counts in his favor we leave to you to decide. When he isn't in his office grading papers, or at home working on his dilapidated Victorian house, or riding his bike between the two, he's usually in Paris.

Addresses

Mark Siderits
Department of Philosophy
Campus Box 4540
Illinois State University
Normal, IL 61790-4540
Office Phone: (309)438-8602
E-mail: msideri@ilstu.edu
Fax: (309) 438-8028

Department
Department of Philosophy
Campus Box 4540
Illinois State University
Normal, IL 61790-4540
Department Phone: (309) 438-7665
E-mail: Philosophy@ilstu.edu
Fax: (309) 438-8028

