PH1102E/GEK1067

Week 10

The self

I. Why does personal identity matter?

Once you were a small child. Now you are an adult. What is this "you" that was once a child and is now an adult? And whatever it is, exactly in what sense does it survive from one year---or one day, hour, minute, or second---to the next?

These questions lie at the heart of the problem of personal identity. Philosophers have offered a variety of answers to them, several of which I'll discuss below. But before getting into that, let's consider why the questions matter. Why should we care what it is in us that survives over time, and how?

There are at least two reasons.

A. Personal identity and moral responsibility

The first reason relates back to something we considered when we were discussing free will, determinism, and moral responsibility. Recall that some philosophers (such as Galen Strawson) say that determinism and moral responsibility cannot coexist. One way they argue for this is by means of the "Mad Scientist" thought-experiment. In this thought-experiment, a decent, mild-mannered man named Smith falls victim to a mad scientist, in the following way. One night, when Smith is sleeping, the mad scientist---let's call him Dr. Frink---secretly anaesthetizes Smith and plants remote-controlled electrodes in his brain. (Neither Smith nor anyone else ever discovers that Frink has done this.) These implanted electrodes do not give Dr. Frink any direct access to Smith's glands or muscles, but, by activating the electrodes remotely in specific ways, Dr. Frink is able to produce in Smith any desire that he, Frink, wishes Smith to have.

For example, Frink can stimulate in Mr. Smith a desire to murder Mrs. Smith, his beloved wife of 35 years. Since Smith is a decent mild-mannered man, and since he loves his wife, he finds this sudden desire he has to murder her disturbing and alarming. Certainly he is not moved to act on it by murdering his wife; more likely he reacts to it by seeking psychiatric treatment, or psychological therapy.
But Dr. Frink has anticipated this. His electrodes also allow him to shut-down Smith's powers of desire suppression, as well as to erase all of Smith's loving feelings for his wife, eradicate all his fond memories of their time together, and replace all of these with feelings of hatred, and false memories of years of marital strife. Let us suppose that Frink induces all these psychological changes in Smith by remote, and only then stimulates in him the desire to murder his wife; the results for Mrs. Smith are predictably tragic.

According to the the Humean theory of moral responsibility, a person is responsible for what he does provided that he does it intentionally. "But," says the anti-Humean, "in the thought-experiment, Mr. Smith murders Mrs. Smith intentionally. After all, to kill a woman intentionally is just to succeed in doing something that you believe and desire will result in her death. Well, when Mr. Smith stabbed his wife fifteen times with a butcher's knife, he believed and desired that this would result in her death. So,

Smith killed his wife intentionally. Yet it hardly seems right to hold him morally responsible for his wife's murder! And if the judge and jury do find him guilty of murder, that will only be because they do not know about Dr. Frink's electrodes."

But the anti-Humean does not stop there. He goes on to point out that if determinism is true, we are all, in a sense, at the mercy of Frinks. It's just that the Frinks that govern us are not mad scientists, but features of our genetic makeup and environment. We have no more control over these genetic and environmental factors than Smith has over Frink, and these factors determine our psychological traits---beliefs, desires, capacities for desire-suppression, etc.---at least as completely as Frink's activities determined Smith's pyschology. Consistency therefore demands that if we should not hold Smith morally responsible for killing his wife in the scenario described, then neither should we hold anyone responsible for anything---at least, not if determinism is true.

What can a Humean say in response to this? I suggested that he could respond that Mrs. Smith's killer is morally responsible for her death, but that Mr. Smith is not responsible for her death, since Mr. Smith is not Mrs. Smith's killer. What makes this paradoxical-sounding reply possible is that---according the Humean---Dr. Frink's extensive interference with Mr. Smith's beliefs, desires, etc. effectively destroys Mr. Smith, and replaces him with an outwardly indistinguishable, but in fact completely different person. It was this person, who has inherited Mr. Smith's body (thanks to Frink's meddling) who is morally responsible for killing Mrs. Smith. Mr. Smith himself was never at the scene of the crime.[1]

This reply to the anti-Humean relies heavily, indeed entirely, on the claim that the person who married Mrs. Smith was not the same as the person who killed her (their outward bodily resemblance notwithstanding). So here, at the heart of the issue of freedom and responsibility, arises the question of personal identity.

B. Personal identity and the end of life

On some level, most of us fear death. And if you fear death, what you fear is the prospect that you---the "you" that exists now, your self---will someday not exist. Is this fear rational?

Some philosophers argue that it is not rational, on the grounds that your self will always exist. But in order for that to be so, this self must have a very special character---it must, for example, transcend any purely bodily aspect of your nature. The hope for immortality is therefore predicated on some specific conception of the self; it is a hope that is justified only if the "you" that exists now is of such a nature as to be able to buck the second law of thermodynamics indefinitely.

Other philosophers argue that the fear of death is irrational, but for completely different reasons. These philosophers believe that it is irrational to fear death, even if death is inevitable---even if your self will perish along with your body. According to these philosophers, if you attain a proper understanding of what your self really is, and of what it really means for it to survive from one day to the next, you will realize that your instinctive attachment to your self is a mere prejudice. Once you understand what the

[1] If you find yourself inclined to say that the person who killed Mrs. Smith shouldn't be held accountable no matter who he is, since, whoever he is, he did what he did only because of the bad character he received from Frink, you should reflect that we usually do not absolve child-abusers of moral responsibility for their abusive behavior, even though in many cases this behavior comes of a character formed by the abuser's own abusive parents.

persistence of the self really involves, you will realize that you have no more reason to fear your own death than you have to fear the death of anyone else---perhaps less reason.

Finally, there are philosophers who hold that there is nothing irrational about the fear of death. According to these philosophers, the "you" that was once a child and is now an adult will one day be nothing---or at any rate, nothing that can be identified with you. Perhaps the best thing to do in this case is to ignore death, not in the hope that it will go away (it won't), but in order to make the best of what life offers in the interim.

II. The memory theory

The memory theory of the self is based on the idea that you are above all a psychological being: a being with beliefs, desires, and a whole complex of psychological traits that make up what we call "character." Of course, a typical adult is psychologically much different from an infant. The adult "you" has little in common with week-old infant "you," psychologically. Psychologically, the adult "you" more closely resembles an adult chimpanzee than it resembles the infant "you." What makes the infant and the adult the same self---the same "you"---is not, therefore, psychological resemblance or similarity. Rather, it is psychological continuity. For just as your infant body did not develop into your adult body in a single sudden leap, neither did your infant mind develop into your adult mind in a single sudden leap. As you mature, you acquire new beliefs, desires, and character traits gradually---and lose them gradually as well. The psychological transition from infancy to adulthood is profound, but it is a transition that you survive, due to the fact that each stage of your psychological history is an incremental development from the previous stage, which was in turn an incremental development from the stage that preceded it, and so on, and so forth.

Crucially important to this conception of the self is **memory**. Early versions of the memory theory simply held that you---the "you" that exists today---are the same as the you that existed yesterday, because the you that exists today can remember the experiences of the you that existed yesterday. These early theories proved to be too simplistic, however. For example, they implied that you were never a week-old baby, since you---the you who now exists---cannot remember any experiences of a week-old baby.

To address this kind of problem, later versions of the memory theory took a more nuanced approach. According to these theories, in order for a present person B to count as the same person as some past person A, it is unnecessary that B have the ability to remember any experience of A. It is enough if B can remember an experience of a person who can remember an experience of a person who can remember an experience of a person...who can remember an experience of A.

Suppose, for example, that the Richard can remember his adventures as a Dick, but cannot remember anything that he did as little Richie. Richard still counts as the same person as Richie, provided that Dick could remember the experiences of Richie. In this case, we can say that Richard is psychologically connected with Dick, and that DIck is psychologically connected with Richie, but that Richard is not psychologically connected with Richie. (At least, he is not psychologically connected with Richie in terms of memory.) Still, there is an indirect psychological link between Richard and Richie---a link mediated by the connection between Richard and Dick, and between Dick and Richie. We can call this indirect link "psychological continuity," and say that Richard and Richie are psychologically continuous with one another, despite the fact that they are not psychologically *connected* to one another.

The memory theory is by far the most popular account of personal identity. It exists in many variations, but the central idea is always the same: for an earlier self A to be the same self as a later self B, is for A and B to be *psychologically continuous* with one another, either via memory or some other psychological trait or traits.

III. Objections to the Memory Theory

The (refined) memory theory is very popular, but it still faces some important objections.

A. Senility

One objection arises from cases in which a person can remember adventures from his childhood, but cannot remember anything from his life as a young adult. (This is the unfortunate position in which victims of advanced Alzheimer's Disease find themselves.) Suppose that Old Man Richard is such a person. Old Man Richard can remember doing things as little Richie, but he can't remember anything about his more recent life, up to this morning's breakfast; among other things, he can't remember any of Dick's experiences. According to the memory theory, Old Man Richard is the same as little Richie, since Old Man Richard can remember Richie's experiences:

(1) Old Man Richard = Little Richie

Of course, young Dick could also remember doing things as Little Richie, so according to the memory theory, Dick and Richie are the same person:

(2) Little Richie = Dick

But since there is no direct or indirect link between Old Man Richard and Dick (due to Old Man Richard's advanced Alzheimer's), the memory theory tells us that Old Man Richard and Dick are not the same person:

(3) Old Man Richard ≠ Dick

But wait! If A = B and B = C, then A = C. So since the memory theorist asserts (1) and (2), the memory theorist must also assert that Old Man Richard = Dick. But the memory theorist also asserts that Old Man Richard ≠ Dick. The memory theorist is caught in a contradiction!

To solve this problem, a memory theorist can modify his theory. He can say that a person, P1, existing at one time is the same as a person, P2, existing at an earlier time if and only if either P1 is psychologically continuous with P2 (in the sense explained above), or there is some person, P3, whose experiences both P1 and P2 can remember. Since both Dick and Old Man Richard can remember Little Richie's experiences, this modified version of the memory theory doesn't assert (3): it asserts that Old Man Richard and Dick are the same person.

B. Amnesia

A different objection to the memory theory is that it equates the onset of total amnesia with death. Suppose that Richard bumps his head while playing football, and consequently loses all of his memories. From that point on, he can't remember anything that happened before that fateful bump. His

personality is unaffected, and he retains all his general mental abilities (intelligence, linguistic competence, etc.). It's just that his episodic memories are completely destroyed.

According to the memory theorist, Richard does not survive the bump. After all, after the bump, there is no one who can remember any of the experiences that Richard had, and no one who is psychologically continuous with any stage of Richard's life (from birth up to the bump). The person who exists after the bump looks like Richard, and sounds like Richard, and has the same personality as Richard, but he is not, according to the memory theorist, Richard.

To this, memory theorists respond that their theory is getting this case right. Richard doesn't survive the bump. The amnesia ends one self, and begins a new self. The new self and the old self are similar in various respects, but that doesn't change the fact that they are different selves (and different people). Or so say the memory theorists.

C. Surgical paradox

Perhaps the most serious objection to the memory theory takes the form of a paradox that we can illustrate with the following example.

Suppose that you are captured by mad but highly-skilled scientists. The scientists clone two copies of your body, minus the brain; call these brainless body-clones Alpha and Beta. The scientists then remove your brain from your (original) body, and implant half of it into Alpha, and the other half into Beta.

Now, it turns out that you can survive with just half a brain; this is what makes the procedure known as a hemispherectomy possible (you read about this procedure in the assigned reading). And it doesn't matter which half you are left with: you can survive the loss of either half (though not, of course, the loss of both halves).

By the same token, when Alpha is supplied with one half of your brain, Alpha becomes a conscious, thinking being---a self.

Likewise for Beta.

So, question: after all of these surgical shenanigans, who are you? There seem to be only four options to consider:

(1) You are Alpha, but not Beta.
(2) You are Beta, but not Alpha.
(3) You are both Alpha and Beta.
(4) You are neither Alpha nor Beta.

We can rule-out options (1) and (2) immediately. That's because for any reason you can give in favor of equating your post-surgical self with Alpha, there is an equally good reason you could give in favor of equating your post-surgical self with Beta, and for any reason you can give in favor of equating your post-surgical self with Beta, there is an equally good reason you could give in favor of equating your post-surgical self with Alpha.

How about (3)? This is hard to accept. After the surgery, Alpha and Beta have distinct thoughts and conscious experiences, they can have conversation or an argument with one another, they might go their separate ways, raise separate families, never see one another again. But this should be impossible if you are both Alpha and Beta, since in that case Alpha and Beta are just one person (namely, you).

This leaves us with (4). But this option too looks unacceptable. If you take option (4), you're saying that you don't survive the procedure at all. But remember: hemispherectomy patients do (apparently) survive the loss of one-half of their brains. If the surgeons had only transplanted half your brain, and discarded the other half, they would, in effect, have performed a hemispherectomy on you. Since hemispherectomy is survivable, you would have survived if they had transplanted just one hemisphere. So how can the fact that they transplanted more than one hemisphere result in your non-survival? You can also think of it this way: if option (4) is correct, then neither Alpha nor Beta can be sure that he ever had a whole brain, until he learns whether the other half of the brain (the one that is not in his head) was saved or discarded. For example, Alpha will have to say that he has no personal history (as a whole-brained individual) unless the surgeons destroyed the other hemisphere (the one to which Alpha's own hemisphere used to be attached).

How does this relate to the memory theory?

Well, both Alpha and Beta can remember having experiences as the pre-surgery you. So, according to the memory theory, both Alpha and Beta are you. You are Alpha, and you are Beta:

(1) Alpha = you

(2) Beta = you

But if A = B and C = B, then A = C. So, the memory theorist has to say

(3) Alpha = Beta

But Alpha is quite clearly not identical to Beta! They might be similar to one another, but they are distinct centers of thought and consciousness---distinct selves. So it looks like the memory theorist is once again caught in a contradiction.

To solve this problem, memory theorists (like Derek Parfit) argue that we should stop thinking about persons and selves in terms of identity, and think about them instead in terms of survival, where we take survival to be something that comes in degrees.

So, instead of saying that you are identical with Alpha (i.e., (1)), we should say that you survive to a certain extent as Alpha; and, instead of saying that you are identical with Beta (i.e., (2)), we should say that you survive to a certain extent as Beta. From these survival-claims we cannot infer that Alpha is the same as Beta (or that Alpha survives as Beta), thus avoiding the contradiction.

IV. Some radical proposals

Many people find the memory theory unsatisfying. Some find it unsatisfying because it forces us to accept that personal identity (or rather: personal survival) comes in degrees. Some find it unsatisfying because it doesn't assign enough importance to consciousness (as opposed to psychological traits like

memory). You can learn more about these criticisms of the memory theory in more advanced philosophy modules.

Here, I want to focus on two radical proposals. One of them rejects the memory theory and indeed every other theory of the self. The other offers the memory theorist a way out of the surgical paradox, without abondoning the idea of identity (in favor of degreewise survival).

A. No self?

According to the classical Indian philosopher Nagarjuna, the key to solving the surgical paradox is to recognize that there is no such thing as the self. On Nagarjuna's view, the self is an illusion---an almost inevitable illusion, but an illusion all the same.

On this view, the right way to see the surgical paradox is as proof that selves do not exist! The proof goes like this:

P1. If you have a self, then after the surgery, one of options (1), (2), (3), and (4) must be correct.
P2. But (as argued above) none of these options is correct.
P3. Therefore, you do not have a self. (follows from P1 and P2)

Needless to say, the idea that you do not have a self is hard to grasp. Still, it is an idea to which many people have been drawn, and not just within the Buddhist tradition. For example, David Hume famously (or infamously) argued that self was an illusion. You can learn about his argument if you take one of our modules devoted to Hume's philosophy.

B. Multiple selves per body?

One way to overcome the surgical paradox is by denying that there are any selves at all. Another way is by asserting that a normal human body contains two selves: one self per brain-hemisphere.

How does this overcome the paradox? Well, if there were two selves in your body even before the surgery---call them Left Brain Self and Right Brain Self---then the word "you" as it occurs in the paradox (and in the options (1) through (4) discussed earlier) is ambiguous: it could refer either to Left Brain Self or to Right Brain Self. If it refers to Left Brain Self, then the correct option to take is (1) (assuming that Alpha receives the left hemisphere). If it refers to Right Brain Self, then the correct option is (2) (assuming that Beta receives the right hemisphere).

The downside to this solution, of course, is that it is hard to believe that each of us is actually two people---two conscious, thinking beings---packed into a single body. When I say "I," who does this "I" refer to? Am I the Left Brain Self or the Right Brain Self? And whichever one I am, why do I get along so well with the other self that also inhabits this body?

Let's return to the memory theory. If we the multiple-selves account is correct, the memory theory has an easy solution to the surgical paradox. Before the surgery, there are two selves in one body. After the surgery, there are two selves in two bodies. Each of the post-surgical selves is the same as the pre-surgical self whose memories the post-surgical self can remember. By this account, the brain-surgery is not fundamentally different from a procedure that separates conjoined twins.

V. Conclusion

There is presently no consensus among philosophers about what the correct theory of the self is, or even about whether there are such things as selves. In this discussion we have only scratched the surface of the debate. (For example, we haven't even considered theories that equate selves with organic bodies, or theories that give consciousness priority over cognitive features like memory.) One thing, however, seems certain: whatever the correct theory of the self turns out to be, it is unlikely to corroborate all of our ordinary everyday beliefs about our own innermost natures. Whatever the self is (if it exists at all!), its nature is almost certainly not what you now take it to be.