

## Summary of Parfit's Reasons and Persons

What makes you, dear reader, the same person who fell asleep in your bed last night? If you believe in a soul, then it might be that you have the same soul as the person who got out of your bed. But we have little reason to believe in souls. It can't be the continuity of consciousness, because you were the same person last night too, before you regained consciousness this morning. So what is it?

You are not who you believe yourself to be. What most of us believe is common-sense about morality makes each of us worse off. Climate change policy doesn't benefit anyone. These are each implications of the central themes in Derek Parfit's famous monograph "Reasons and Persons". Parfit's book is about ethics, about what we have reason to do, not the more abstract philosophical questions about what we can know about ethics, or whether ethical statements can be objectively true or false. In arguing for his positions, Parfit makes extensive use of science fiction analogies to work our intuitions about morality. In this note, I summarize his main points with an eye towards my economics colleagues.

### Personal identity

If we do not have souls, then it must be either physical continuity or psychological connectedness that makes us the same person. Physical continuity is the idea that it is having the same brain that makes us the same person over time, however our personality or memory changes. Psychological connectedness claims that it is having a large enough set of overlapping memories and personality traits makes us the same person. Parfit uses examples to show that both of these ideas are problematic. First, the problem with psychological continuity or connectedness making us the same person.

### Teletransporting

The year is 2100. Teletransporting technology has largely replaced flying as a means of travel. Teletransporting machines make a detailed scan of one's body, rebuild the body in a different location, and destroy the original. For many years, you avoided travelling long distances because teletransport terrified you. Isn't having your body destroyed a kind of death? Your spouse who frequently teletransports as part of his job makes fun of you for being so worried. He is fine, you will be too.

Then you had a scheduled work trip to the new colony on Mars. The only practical way of getting to the colony was teletransport. You overcame your fear, got into the machine, and pushed the big green button. There was a whirring sound, but to your great relief after a few moments the door unlocked and you stepped out onto Mars. Since that trip you have frequently teletransported, and it has become routine.

So far, this story, and at least my intuition suggest it isn't physical continuity

that makes a person the same person over time. In the teletransporter, it doesn't matter that your physical body has been destroyed and an exact replica has been created many times. What matters is psychological continuity and connectedness. Being teletransported wasn't a way of dying at all, it is simply a way of moving to a new location quickly.

One day, after entering the teletransporting machine, you push the green button and hear a whir, but when the door unlocks you are still where you started. A scientist approaches and tell you that they have been installing some new equipment, and something went wrong. The replica of your body was successfully created on Mars as usual, but the machine did not destroy your original body. Moreover, there is some bad news. Even though it didn't destroy your body, the radiation from the machine has seriously affected your heart. Your heart will cease functioning in the next 24 hours, and there is nothing to be done about it. The only good news is that death will be totally painless and instant.

Should you be distressed about your heart failing? The version of you on Mars has all of your memories and your personality. It loves your family as much as you do, and will work toward carrying out all your intentions just as you would have. You have, after all, teletransported many times before. If your body had been destroyed as usual, you would have considered the replica to be you. By this argument, maybe you should not feel bad about your impending heart failure.

On the other hand, there cannot be two yous. Personal identity is by definition singular. Since you have both psychological connectedness and physical continuity with the person who walked into the teletransporting machine, surely you have the better claim to being the true you. The person on Mars is merely an exact replica. Intuitively, if I were in your shoes, I would feel quite anxious about my impending death.

This example challenges the claim that psychological continuity and connectedness is what matters. An alternative claim is that physical continuity is what matters. On this view, teletransporting is a way of dying. Parfit gives a different example to call physical continuity into question.

### **The mad scientist**

In 2100, you never take the teletransporter. As a believer in the physical continuity criterion of personal identity, you believe that teletransporting is dying. Unbeknownst to you, a mad scientist has taken a keen interest in you. Every night while you sleep, he sneaks into your house and replaces a few of your cells with identical copies. After one year, none of the original cells remain. According to your views, the person living your life is no longer you. It is merely a replica exactly like you.

The question is when you died. Certainly on the first night, when the mad scientist replace a few of your cells, this was still you. On the last night, when

nearly all of your cells had been replaced, this was certainly not you according to your physical criterion. There must have been one night when you died. The previous day it was you, but then next it was someone else. But then the difference between being alive and being dead was only a small number of cells. How could replacing only a few cells with identical copies be a matter of life and death?

## **Conclusions on personal identity**

Parfit ends up concluding that personal identity is not what matters. What matters is psychological continuity and connectedness. We can still talk about personal identity, but it will just be a convention. We can call teletransportation dying, or not. What is important is that you will have a strong psychological connection with your replica. You might even have many replicas, and have strong psychological connections with each of them. Whether they are you is an empty question.

If it is psychological connectedness that matters, as we live our lives we have less psychological connections with our childhood selves. Our distant future selves are also less connected to our present selves than we might have imagined. We might even be connected to other people, like our spouse or close friends, than we are to our childhood selves. This fact may well provide a reason for us to discount our future, which is not directly related to time itself. We discount the well-being of our future selves because they are less psychologically connected to our present selves, and that is what matters.

## **Common-sense morality**

Most people believe that we have special obligations to ourselves, our children, our students, our clients, and our compatriots. We should do what is best for our children, even if this is somewhat worse for other people's children. We should do what we can to help our students find satisfying employment, even if our efforts make things harder for other students. And so on.

This is what Parfit calls common-sense morality. He argues that common-sense morality is collectively self-defeating. That is, if everyone follows common-sense morality, its goals are worse satisfied than if everyone did something else. For students of economics, this point will be familiar, as it is simply the standard outcome of non-cooperative games such as prisoners dilemmas or public good provision problems.

For example, we all might prefer to have our children have low-stress childhoods, with plenty of time to pursue their own interests. If they are a little more prepared for college entrance exams than their peers, though, they will be more likely to get into the most prestigious programs and get a leg up in the labor market. Therefore we send our children to college prep courses. But if all children go to college prep courses (as is true in Taiwan and South Korea), then

no one gets any advantage, and the children have no free time and high-stress childhoods.

Should we reject common-sense morality because it is self-defeating? Parfit argues that we should. A moral code is fundamentally about how people should all act. For example, Kant believes that we should follow a moral rule if we believe it should be a universal law. If by all successfully following a theory, its goals are worse achieved, then the theory should be rejected.

The reason that common-sense morality is self-defeating is that it is agent-relative. Each person has different goals, and when each pursue them individually, their goals are less well satisfied than if they had cooperated. Other types of morality, for example utilitarianism, are agent-neutral. Every person has the same goals. Since there is no conflict of goals, agent-neutral morality is not collectively self-defeating. Parfit argues we that we must revise common-sense morality to make it agent-neutral in situations where it is self-defeating.

## Future generations

According to Judith Harris's summary of personality research, 50% of the variation in personalities is due to genes. Genes also determine how we look, and though we might resemble our siblings, unless we are identical twins we do not look the same as them. It is natural to conclude that if someone had a substantially different genetic makeup, he would not be me. While each of us shares half of our genes with each parent, exactly which genes depends on which ovum and sperm cell fused to form our first cell. In each ejaculation, a healthy male releases 100s of millions of sperm cells. We are each one in a hundred million.

Since the births of particular people are so unlikely, even small changes in policy are likely to completely change the set of future people who will live in, say, two hundred years. Suppose that we take drastic action to arrest climate change, or at least to insure against the worst disaster scenarios. Then the set of people who will live in the future are not the same as who would live if we were to continue with the status quo.

One possibility is that the same number of people might live in the climate change action scenario as in the status quo. It seems natural that all else equal we should prefer the climate change action scenario, because then the future people have better lives. Parfit calls this the *principle of beneficence*, that if the number of people are the same, we should prefer the scenario in which people have better lives.

The rub comes if we consider two scenarios with different numbers of people. Suppose that in one scenario, there are a small number of people with lives much better than those of people alive today. In another, there are more people with lives better than people alive today, but not much better. If there are enough

people in the second scenario, then it is natural to prefer it to the first scenario. But if we can trade off quantity and quality, we could imagine even more people with lives about the same as people alive today would be preferable to the second scenario. And then it would be even more desirable if we had yet more people with lives not as good as people alive today, but still worth living. We can keep iterating on the process until we conclude that vast multitudes of creatures with lives only barely worth living, maybe insects on some primitive planet, would be preferable to a planet like Earth but where each human life is much better than current people's lives.

Parfit calls this *the repugnant conclusion*. He shows that is impossible to avoid with utilitarianism and other forms of consequentialism. Parfit spends the last section of the book searching for a theory which both implies the principle of beneficence, but avoids the repugnant conclusion. He does not find a satisfactory theory, but ends the book on the hopeful note that someone else will be able to succeed where he has failed, now that the problem has been identified. This was the beginning of the philosophical field known as *population ethics*.

## Read it yourself

*Reasons and Persons* is a difficult book. Parfit's audience is the professional academic philosopher, which leads him to be very careful with precise phrasing. This makes the writing dense, and often full of double negatives.<sup>1</sup>

Furthermore, many of the chapters address academic critiques which I believe are of little interest to the general reader. If you are intrigued by part of the summary I wrote, but do not have the time or interest to read the entire manuscript, I single out self-contained parts of the book to read below.

### Rejecting common-sense morality

- Chapter 1.1 (just the very beginning)

This defines self-interest theory, which will be used through the first two parts of the book.

- Chapter 2

All about public goods and other game theory that economists will be familiar with.

- Chapter 4

---

<sup>1</sup>Consider this paragraph: Some may claim that our choice of Depletion does not have a bad effect. This cannot be claimed about our choice of the Risky Policy. Since this choice causes a catastrophe, it clearly has a bad effect. But our choice will not be bad for, or worse for, any of the people who later live. This case forces us to reject the view that a choice cannot have a bad effect if this choice will be bad for no one.

This is the section where Parfit argues that we should reject common-sense morality

### **Personal identity**

- Chapter 10, Chapter 11

In these Chapters, Parfit introduces a number of thought experiments which challenge our intuitions about personal identity.

- Chapter 12, Chapter 13

Here is where Parfit argues that personal identity is just a convention, but should not carry any moral weight. He describes what does matter, psychological continuity and connectedness.

### **Future generations**

- Chapter 16

This is the description of how policy changes affect the identity of future people

- Chapter 17

This chapter is about the repugnant conclusion