For example, following a rule such as "Do not kill healthy patients" may have worse consequences in a particular instance such as the transplant case (four people will die, as opposed to a single person), but it may be that adopting it as a general rule has the best consequences overall (the hospital system will work better and many more people will survive). But what about a qualified rule that says, "Kill someone when this will save more lives and no one will ever know"? Even if adopting some rule like this would have better consequences overall, it can still seem immoral. So there remain some puzzles here.

In her classic 1958 article "Modern Moral Philosophy," Elizabeth Anscombe was scathingly critical of both consequentialist and deontological approaches to ethics. She said that consequentialism leads to immoral consequences and reveals a "corrupt mind." On the other hand, deontological approaches such as Kant's involve a legalistic conception of morality as a set of laws laid down by a legislator. Anscombe thinks this picture is a leftover from the old divine command theory, according to which God was the legislator. For Anscombe, once God is rejected, the rule-governed approach cannot work.

Anscombe thought we should not talk of what's morally right and wrong at all. These terms are too coarse to capture what is of interest in morality. Instead, we should use finer-grained words like "unjust" and "brave" and "kind" to assess the moral character of people's actions.

In this, Anscombe was recommending a return to the *virtue ethics* found in Aristotle, in which ethics centers on virtues such as bravery and kindness. A closely related picture is found in the work of Confucius, Mencius, and other philosophers in the Confucian tradition, which gives a central role to the moral traits we should aspire to, such as benevolence, trustworthiness, and wisdom. One popular version of virtue ethics frames the moral character of an action in terms of the moral character of a person who would perform that action. A brave act is one that a brave person would perform. A kind act is one that a kind person would perform.

Thanks to Anscombe as well as her Oxford colleagues Philippa Foot and Iris Murdoch and Chinese philosophers in the New Confucian movement, among others, virtue ethics has recently had a resurgence as a leading moral theory. It is sometimes criticized for not giving clear criteria regarding how we should act. Nevertheless, it's often seen as offering tools for moral improvement and for understanding the rich tapestry of morality without reducing it to simple principles.

Simulations and moral status

Let's turn to the ethics of virtual reality. We can start by thinking about long-term simulation technology. When is it morally permissible to create a simulation? When is it morally permissible to end one? What are our moral responsibilities as creators of a simulation?

If we simulate a universe without life, there are few ethical issues. Cosmologists already run simulations of the history of galaxies and stars, and they don't need permission from an ethics board. Perhaps there are ethical issues about whether this is the best use of computer power, and about what to do with knowledge gained from the simulation, but these are standard ethical issues involved in everyday science. Even simulating biology—say, at the level of the evolution of plant life—does not go far beyond this.

Ethical issues arise when we simulate *minds*. To start with an extreme case, say that we work for an intelligence agency and we want to simulate the reactions of human beings who undergo torture. We create simulated humans with fully functioning simulated brains and subject them to (simulated) torture. Is this morally acceptable or morally horrific?

A natural reaction is that it depends on the mental life of the simulated beings. If they're conscious creatures who experience suffering, simulated torture would be morally horrific. If they're unconscious simulations and don't experience suffering, then simulated torture would perhaps be morally acceptable.

This raises a fundamental issue: Do sims have moral status? A being has moral status when it's an object of moral concern in roughly the way that people are objects of moral concern—that is, it's a being whose welfare we need to take into account in our moral deliberations.

A being has moral status when that being *matters*, morally speaking. The Black Lives Matter movement is all about moral status. Black lives matter as much as any human lives do! Killing Black people is as bad as killing white people. Mistreating Black people is as bad as mistreating white people. In the past, and even today, many people and many social institutions have treated Black lives as if they mattered less than white lives. This is now widely recognized as monstrous.

Over the years, the circle of moral status has expanded. It's now widely accepted that many nonhuman animals have moral status, too. The issue isn't quite the same as with human lives. Most people think that humans matter more than birds and dogs—but birds and dogs still matter to some extent. We shouldn't be wantonly cruel to dogs. It's less obvious whether flies and shellfish have moral status; some people think they do. Some environmentalists hold that trees and plants have a sort of moral status, but this is a minority view. As for inorganic matter, few people think that rocks or particles have moral status. You can treat a rock however you like, and this won't matter morally, at least as far as the rock is concerned.

My own view (shared with many others) is that what bestows moral status is *consciousness*. If an entity has no capacity for consciousness, and never will have, then it has no moral status. It can be treated as an object. If an entity has the capacity for consciousness, then it has at least some minimal moral status. If it can experience something, that should be taken into account in our moral calculations. It's arguable that systems with a minimal degree of consciousness (ants?) have only a minimal degree of moral status and so weigh much less heavily than humans in our moral deliberations. But consciousness at least gets them in the door.

We can use a thought experiment to help us think about the moral status of consciousness. I call it the *zombie trolley problem*. You're at the wheel of a runaway trolley. If you do nothing, it will kill a single conscious human, who is on the tracks in front of you. If you switch tracks, it will kill five nonconscious zombies. What should you do?

A few clarifications. The zombies are philosophical zombies, as described in chapter 15: near-duplicates of human beings with no conscious inner life at all. Zombies have no subjective experience. You can imagine them as physical duplicates of us without consciousness, or as silicon versions of us without consciousness if that's easier. If that's still too hard, imagine something as close to us as possible without the capacity for consciousness. Whether or not these zombies will be useful for various purposes isn't relevant in this thought experiment; what matters is their moral status.

When I have taken polls about the zombie trolley problem, the results are pretty clear: Most people think you should switch tracks and kill the zombies. It's worse to kill one human than five zombies. A few say that zombies count for as much as humans, so we should kill the human, but they are a distinct minority.

Killing the zombies may sound awful. There is a recent movie, *Zombies*, that centers around the way a community of zombies is mistreated in a human world. But importantly, the zombies in the movie are conscious. Philosophical zombies lack consciousness, so that there is arguably no one home to mistreat.

We can take things further. Suppose you have the choice between killing one conscious chicken or a whole planet of humanoid philosophical zombies. At this point, intuitions are less clear. Some people stick with "Kill the zombies," reflecting the view that without consciousness there's no moral status. Others switch to killing the chicken, presumably because they think the zombies have some degree of moral status, perhaps deriving from their intelligent behavior. My own intuition wavers on this matter.

The zombie trolley problem can lead to a weak or a strong conclusion. If you think a single conscious creature should be saved at the cost of killing five nonconscious creatures, this suggests that consciousness is *relevant* to moral status. Conscious creatures matter more than nonconscious creatures. If you hold the stronger view—that there's never a moral reason to spare nonconscious creatures—this suggests that consciousness is *necessary* for moral status. Nonconscious creatures don't matter at all, morally speaking.

The stronger conclusion dovetails with the view I advocated at the end of the last chapter, that consciousness is the ground of all value.

Whenever anything is good or bad for someone, it's because of their consciousness. Consciousness has value, what a conscious creature values has value, and relations between conscious creatures have value. If a creature has no capacity for consciousness, nothing can be good or bad for it from its own perspective. And it's natural to conclude that if nothing is good or bad for a creature, then the creature has no moral status.

The view that consciousness is required for moral status is central in discussion of animal welfare. The Australian philosopher Peter Singer, who inspired the contemporary animal-rights movement with his 1975 book *Animal Liberation*, has argued that what he calls *sentience* is what matters for moral status:

> If a being is not capable of suffering, or of experiencing enjoyment or happiness, there is nothing to be taken into account. This is why the limit of sentience (using the term as a convenient, if not strictly accurate, shorthand for the capacity to suffer or experience enjoyment or happiness) is the only defensible boundary of concern for the interests of others.

In ordinary English, "sentience" is roughly equivalent to "consciousness." Singer uses the term more narrowly to describe suffering and the experience of enjoyment and happiness. This is a *kind* of consciousness: Only conscious creatures can suffer or experience enjoyment or happiness. Singer holds that consciousness is necessary but not sufficient for moral status. Not just any sort of consciousness bestows moral status; the conscious experience of positive or negative affective states is required. The same "sentientist" view has been taken by many recent theorists, who hold that the experience of positive or negative affective states is what matters for moral status. This view goes back at least to Jeremy Bentham, who said in the 18th century that where moral status is concerned, suffering is what matters.

I find this view implausible. There's much more to consciousness than the experience of suffering or happiness, and it's not plausible that these other sorts of consciousness don't matter morally. To make this point, we might think about a more extreme version of the unemotional Vulcan Mr. Spock on *Star Trek*.

Let's say that a *Vulcan* is a conscious creature who experiences no happiness, suffering, pleasure, pain, or any other positive or negative affective states. The Vulcans on *Star Trek* aren't quite as extreme as this: they experience lust every seven years and experience at least mild pleasures and pains in between. To avoid confusion with *Star Trek* we could call our version *philosophical Vulcans*, by analogy to philosophical zombies.

As far as I know, no human being is a philosophical Vulcan. There are some reported cases of humans who do not experience pain, fear, or anxiety, but they still experience positive states. A philosophical Vulcan will lack those states as well. They might still have a rich conscious life, with multimodal sensory experiences and a stream of conscious thought about all sorts of complex issues. We've all experienced affectively neutral states in perception and thought. I can see a building or think about a meeting without any positive or negative affect. For a Vulcan, that's what things are like all the time.

Vulcans' lives may be literally joyless, without the pursuit of pleasure or happiness to motivate them. They won't eat at fine restaurants to enjoy the food. But they may nevertheless have serious intellectual and moral goals. They may want to advance science, for example, and

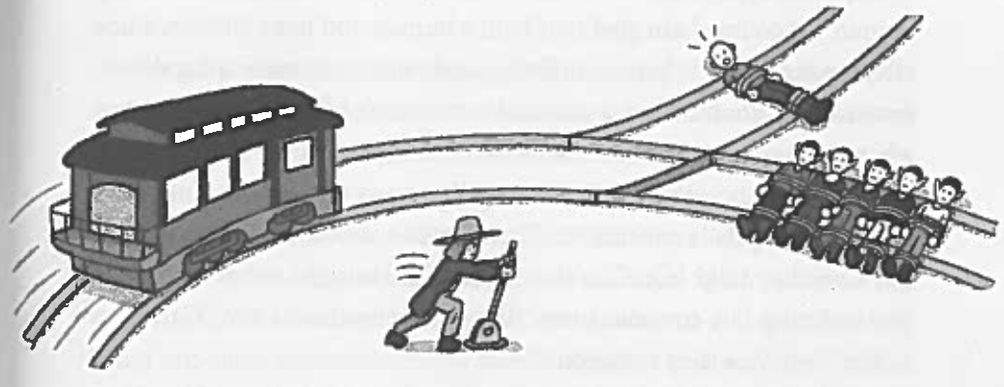

Figure 46 Jeremy Bentham faces the Vulcan trolley problem. Is it better to save one human or five Vulcans?

to help those around them. They might even want to build a family or make money. They experience no pleasure when anticipating or achieving these goals, but they value and pursue the goals all the same.

The Bentham/Singer view predicts that a philosophical Vulcan doesn't matter morally. That seems incorrect. We could make the point with a *Vulcan trolley problem*. Would it be morally acceptable to kill a planet of philosophical Vulcans to save one human with ordinary affective consciousness? I think the answer is obviously no.

More simply, suppose you're faced with a situation in which you can kill a Vulcan in order to save an hour on the way to work. It would obviously be morally wrong to kill the Vulcan. In fact, it would be monstrous. It doesn't matter that the Vulcan has no happiness or suffering in its future. It's a conscious creature with a rich conscious life. It cannot be morally dismissed in the way that we might dismiss a zombie or a rock.

(Does the Vulcan have a desire to keep on living? As I'm thinking of things, yes. If we shouldn't kill such a Vulcan, that shows that more than affective conscious states matter. We could also stipulate an even more extreme Vulcan—one who has no affective conscious states and is also indifferent to continuing to live or dying. My view is that it would also be monstrous to kill this Vulcan. If so, this suggests that more than affective consciousness and desire satisfaction matter. My view is that non-affective consciousness matters, too.)

My own sense is that a Vulcan matters about as much as an ordinary human. Of course I am glad that I am a human and not a Vulcan, since affect makes my life better. Suffering and happiness make a big difference to how good or bad a conscious creature's life is. But they're not what gives a creature moral status in the first place.

Bentham once expressed his view by saying that where the moral status of animals is concerned, "The question is not, Can they reason?, nor Can they talk? but, Can they suffer?" If I'm right, what matters is not suffering but consciousness. The right question is not "Can they suffer?" but "Are they conscious?"

To determine the moral status of simulated creatures, "Are they conscious?" is also the question we need to ask. We've already asked and answered this question for some simulated creatures. In chapter 15, I argued that a perfect simulation of a human brain will be associated with just the same sort of consciousness as the original brain. That is, simulated humans will have the same sort of consciousness as ordinary humans. If consciousness is all that matters for moral status, simulated humans have the same moral status as ordinary humans.

The ethics of simulations

Now we can answer the simulation trolley problem. The answer is no: It is not acceptable to kill five simulated people to save one ordinary human. If simulated humans weren't conscious, this would be acceptable. But because full-scale simulations of humans are conscious in much the same way we are, they have the same moral status as we do.

From a certain perspective, this view may seem unreasonable. Would you really sacrifice a human life in order to save a few computer processes? But we can turn the question around by supposing that we're in a simulation. If we're in a simulation, would it be morally acceptable for our simulators to kill five of us in order to save one of them in the next universe up? From our perspective, I'd say the answer is no. Even if our simulators have the power to do this, this does not make it right. The same goes for our own actions toward conscious people in the simulations we create.

Someone might say that although consciousness matters for moral status, other factors matter, too. For example, maybe nonsimulated humans have higher moral status than simulated humans simply because they're in the top-level universe. Or maybe brief simulations count for less simply because they don't last for as long. I find these views somewhat implausible, though. And once again, considering them under the assumption that we're the ones in a simulation will bring out their downsides. Why should the fact that we're not in the top-level universe make killing us more morally acceptable?