



Right and wrong, good and bad

consequentialism

deontology

the experience machine

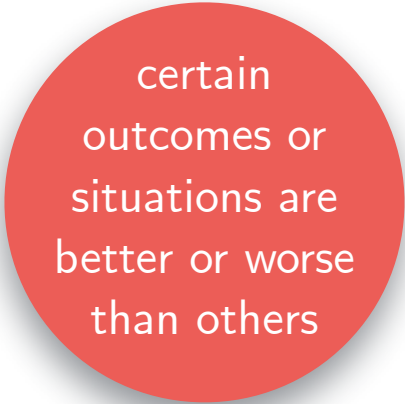
virtue ethics

trolley cases


Last time we talked about the question of whether there are any facts about what is right and wrong or what is good and bad.

Let's suppose, for today, that there are such facts. Well, what are they? What determines what is right or wrong, or good and bad?


Today I want to lay out three general ways of answering this question. To get a grip on these three general ways, it will be useful to distinguish three different kinds of things that we say are good and bad, or right and wrong.



certain
outcomes or
situations are
better or worse
than others



certain actions
are right or
wrong



certain
people or lives
are better or
worse than
others

Our three approaches to thinking about these topics can be distinguished by how they think about the relationships between these facts.

certain
outcomes or
situations are
better or worse
than others

Our three approaches to thinking about these topics can be distinguished by how they think about the relationships between these facts.

Our first view says that the most fundamental of these kinds of facts is the goodness or badness of a situation or state of affairs. Some ways the world can be are just better than others.

What, on this view, does it mean for an action to be right or wrong? Here's a very natural answer:

Consequentialism

An action is right if, of all available options, it would lead to the best outcome, and wrong otherwise.

Similarly, the best people, or the best lives, are the ones whose actions lead to the best overall consequences.

Consequentialism

An action is right if, of all available options, it would lead to the best outcome, and wrong otherwise.

It is easy to construct intuitively powerful arguments in favor of consequentialism. Here is one:

The very simple argument for consequentialism

1. If one has the choice to bring about a worse outcome or a better outcome, one should never choose to bring about the worse outcome.
 2. One should always choose to bring about the best outcome. (1)
-
- C. Consequentialism is true. (2)

Calling this an argument is a bit of a stretch -- the first premise is pretty close to the conclusion. But still: isn't the first premise plausible?

Consequentialism

An action is right if, of all available options, it would lead to the best outcome, and wrong otherwise.

A second argument can be presented via an example.

The drowning child

You are walking to class past St. Joseph Lake, and see a child drowning in the lake. If you don't go in to help the child, the child will die. If you do go in, your clothes will get wet. (You really don't like having wet clothes.) What should you do?

To most people, the answer here seems pretty obvious: you should go in and help the child. But why?

Obvious answer: the situation in which the child lives and you have wet clothes is better than the one in which the child dies and you stay dry. But that assumes that consequentialism is true.

Consequentialism

An action is right if, of all available options, it would lead to the best outcome, and wrong otherwise.

Obvious answer: the situation in which the child lives and you have wet clothes is better than the one in which the child dies and you stay dry. But that assumes that consequentialism is true.

We can present this line of thought as an argument:

The best explanation argument for consequentialism

1. In the example of the drowning child, you should save the child.
 2. The explanation of (1) is that you should always act so as to bring about the best consequences.
-
- C. Consequentialism is true. (2)

Consequentialism

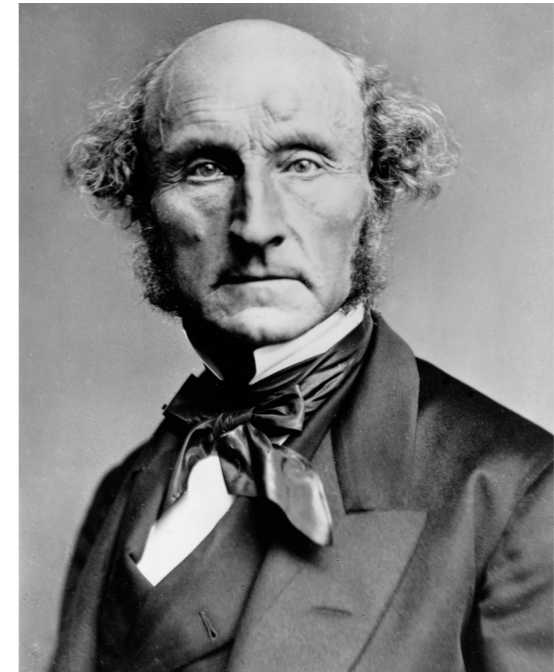
An action is right if, of all available options, it would lead to the best outcome, and wrong otherwise.

Let's suppose for now that consequentialism is true. This raises two questions. The first is: what makes one outcome, or state of affairs, better than another?

To answer this question is to give a **theory of value**: a theory about what makes one state of the world better, or worse than, another.

Let us say that a **good** is something that makes a state of affairs better, and a **bad** is something that makes a state of affairs worse.

In the reading for today, John Stuart Mill gives the following statement of his theory of value — his view of which things are goods and bads.



“The creed which accepts as the foundation of morals ... the Greatest Happiness principle, holds that actions are right in proportion as they tend to produce happiness, wrong as they tend to produce the reverse of happiness. By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure.”

This view is sometimes called hedonism:

Hedonism

Pleasure is the only good and
pain is the only bad.

Suppose that this is true. Then how do we tell whether one outcome is better than another?

Here is a very natural answer. We 'add up' the pleasure, and 'subtract out' the pain. Whatever situation has the highest 'net pleasure' is the best. This gives us the following view:

Utilitarianism

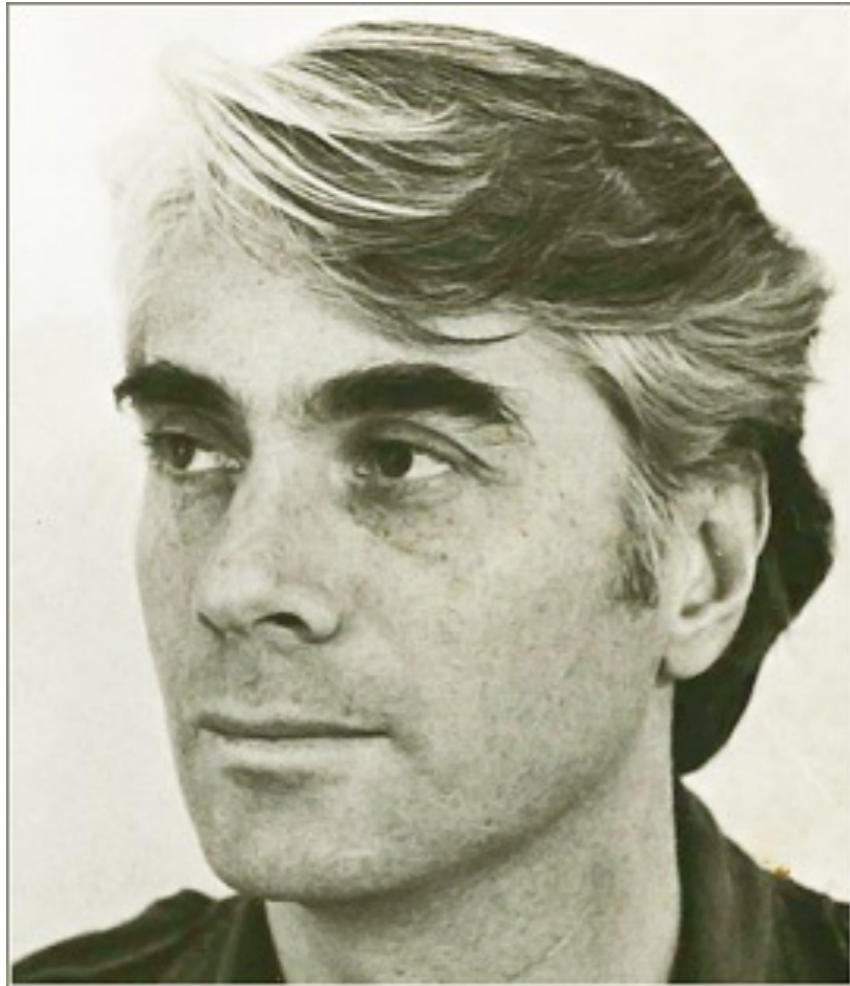
An action is the right thing to do in certain circumstances if, of all the actions available in those circumstances, it would produce the highest net pleasure.

Utilitarianism

An action is the right thing to do in certain circumstances if, of all the actions available in those circumstances, it would produce the highest net pleasure.

Utilitarianism is a very clear and plausible-sounding view about ethics. This is the view which is often summed up with the slogan that one ought always to act to cause the greatest happiness for the greatest number. It is a paradigmatically unselfish theory: no one's pleasures and pains are more important than anyone else's.

A serious challenge to utilitarianism can be brought out by Robert Nozick's example of the experience machine.



“Suppose there were an experience machine that would give you any experience you desired. Superduper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank, with electrodes attached to your brain. ... Would you plug in? *What else can matter to us, other than how our lives feel from the inside?*”

What must the utilitarian say about the relative goodness of the state of affairs in which everyone (or almost everyone) plugs in and the state of affairs in which no one does?

Suppose you face the decision whether to get into the experience machine. What would a utilitarian say about what you ought to do?

Suppose now that you face the decision of whether you should put **everyone** into the experience machine. (The machines are maintained by extremely reliable robots.) What would a utilitarian say about what you ought to do?

Does it matter if people ask you (or beg you) not to plug them in?

Is Nozick right that these consequences of utilitarianism are incorrect?

Recall that we presented Utilitarianism as the combination of hedonism with consequentialism.

You might think that the experience machine is a problem for hedonism, but not for consequentialism as such. Couldn't the consequentialist just say that there are goods besides pleasure, and bads besides pain?

Here are some other candidates for goods:

The extent to which
the desires of
agents are
satisfied.

The extent to which
the states of affairs
contain beauty, or
love, or friendship,
or something else
taken to be of
objective value.

The extent to which
the states of affairs
maximize the well-
being, or welfare, of
agents.

Corresponding to each of these views about the good is a different version of consequentialism. For example, the first would yield the result that one should always act in such a way that maximizes the number of desires of people which are satisfied.

What would that view say about the experience machine?

Our last topic on consequentialism is an attempt to show, not that some specific form of consequentialism fails, but rather that this basic thesis is false, and so that any consequentialist approach to morality should be rejected.

One general feature of consequentialism is its indifference to **how** consequences are brought about. What matters when deciding what to do is what one's various actions will bring about, not what those actions are.

One consequence of this general feature might be stated like this:

Act/omission indifference

Whether I bring about some state of affairs by doing something or failing to do it is morally irrelevant.

Some aspects of this principle are quite appealing. For example, the principle refuses to let people stand idly by as others suffer, on the grounds that one is not the **cause** of that suffering. One whose failure to act leads to suffering is, according to consequentialism, just as responsible for it as one whose action leads to that suffering.

But some troubling consequences of this principle are brought out by the following example, due to Judith Jarvis Thomson.



David is a great transplant surgeon. Five of his patients need new parts—one needs a heart, the others need, respectively, liver, stomach, spleen, and spinal cord—but all are of the same, relatively rare, blood-type. By chance, David learns of a healthy specimen with that very blood-type. David can take the healthy specimen's parts, killing him, and install them in his patients, saving them. Or he can refrain from taking the healthy specimen's parts, letting his patients die.

What does the consequentialist say that David ought to do in this case?
What ought he to do?

This sort of case might lead you to think something like this: killing someone in order to save the lives of others is never morally permissible.

If this were true, this looks like it would be trouble for the Consequentialist, since it is hard to argue that killing someone, especially when it saves the lives of others, can never lead to an outcome which is, overall, the best of the available options.

But, as some of Thomson's other examples show, matters are not quite this simple.

Consider one of her examples involving a trolley car:

Edward is the driver of a trolley, whose brakes have just failed. On the track ahead of him are five people; the banks are so steep that they will not be able to get off the track in time. The track has a spur leading off to the right, and Edward can turn the trolley onto it. Unfortunately there is one person on the right-hand track. Edward can turn the trolley, killing the one; or he can refrain from turning the trolley, killing the five.

Is it permissible for Edward to turn the trolley? If so, wouldn't this be a case in which it is permissible — perhaps even obligatory — to kill one person in order to save five lives?

But then why might it be OK for Edward to turn the trolley, but clearly not permissible for the doctor to cut up his healthy specimen?

One might try to explain the difference here like this: Edward is choosing between **killing one and killing five**; either way, he is killing someone. David is choosing between **killing one and letting five die**, and this is something quite different. We have a stronger duty to avoid killing than to prevent people from dying.

But it is not clear that this is the right explanation of the difference between Edward and David, as is brought out by the example of Frank.

Frank is a passenger on a trolley whose driver has just shouted that the trolley's brakes have failed, and who then died of the shock. On the track ahead are five people; the banks are so steep that they will not be able to get off the track in time. The track has a spur leading off to the right, and Frank can turn the trolley onto it. Unfortunately there is one person on the right-hand track. Frank can turn the trolley, killing the one; or he can refrain from turning the trolley, letting the five die.

Here it seems as though Frank is faced with a choice between letting five die, and killing one — so his choice seems, in this respect, just like David's (the surgeon's). But it seems as though it **is** morally permissible for Frank to turn the trolley, even though it is **not** morally permissible for David to cut up the healthy specimen.

More complications are introduced by yet a third trolley example:

George is on a footbridge over the trolley tracks. He knows trolleys, and can see that the one approaching the bridge is out of control. On the track back of the bridge there are five people; the banks are so steep that they will not be able to get off the track in time. George knows that the only way to stop an out-of-control trolley is to drop a very heavy weight into its path. But the only available, sufficiently heavy weight is a fat man, also watching the trolley from the footbridge. George can shove the fat man onto the track in the path of the trolley, killing the fat man; or he can refrain from doing this, letting the five die.

Many people think that it is not permissible for George to push the fat man. But why is this any different from turning the trolley to kill the one on the right hand section of the trolley tracks? After all, in both cases, you are killing one rather than letting 5 die.

One thought is this: the fat man has a **right** not to be pushed onto the tracks in a way that people standing on trolley tracks **don't have a right** not to be run over by trolleys.

This sort of thought also promises to make sense of the example of David the surgeon; perhaps healthy specimens have a right not to be cut up, but that dying patients in need of transplants have no right to be saved.

This way of thinking about these cases is very different than the way of approaching them suggested by Consequentialism. According to this view, we should think about what we ought to do by **first thinking about the rights and obligations of the people involved** and not, at least in the first instance, about which action would bring about the best outcome.

If you think we should go in some direction of this kind, you should think about how best to reply to the simple but strong arguments for consequentialism with which we began.


certain
outcomes or
situations are
better or worse
than others

certain actions
are right or
wrong

certain
people or lives
are better or
worse than
others

We began with three different kinds of things to which we apply the categories of good and bad, or right and wrong.

The consequentialist takes the goodness or badness of situations as fundamental, and tries to explain the others in terms of these. What would it look like if we took one of the other categories as most fundamental?



certain actions
are right or
wrong

The consequentialist takes the goodness or badness of situations as fundamental, and tries to explain the others in terms of these. What would it look like if we took one of the other categories as most fundamental?

A **deontological** approach to what we ought to do takes facts about the rightness and wrongness of actions as basic, and does not try to explain the rightness or wrongness of actions in terms of the goodness or badness of the action's outcomes.

There are lots of different deontological theories. I'm going to briefly introduce you to one, which is based on the idea that people have certain basic **rights**. On this view, wrong actions are actions which violate someone's rights.

But what are our rights, and which actions violate them?

There are lots of different deontological theories. I'm going to briefly introduce you to one, which is based on the idea that people have certain basic **rights**. On this view, wrong actions are actions which violate someone's rights.

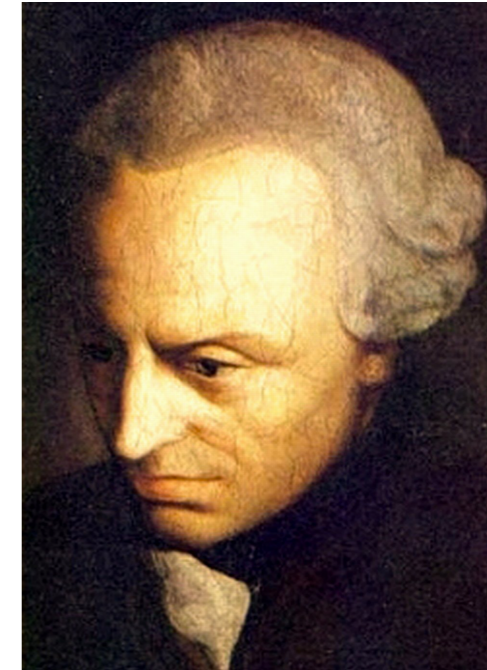
But what are our rights, and which actions violate them?

One answer to this question was given by the 18th century German philosopher Immanuel Kant.

Kant thought that all human beings have dignity, and that this means that we all have the right not to be **used** as a way to satisfy the desires of others.

He called this view 'the law of humanity':

'Act in such a way that you always treat humanity, whether in your own person or the person of any other, never simply as a means, but always at the same time as an end.'



He called this view 'the law of humanity':

'Act in such a way that you always treat humanity, whether in your own person or the person of any other, never simply as a means, but always at the same time as an end.'

What does it mean to use someone as a means to an end, versus treating them as an end in themselves? Think of the complaint that someone is simply using you. When we say this, we are saying that the person is not taking you into account; that he is treating you as a vehicle for his own ends, rather than as deserving respect and consideration in your own right. This is treating someone as a mere means rather than as an end in himself.

That said, it is important to see that the formula of humanity does not prohibit using someone as a means to an end, but only doing so without also treating them as an end in themselves. When you order food at a restaurant you are treating the person to whom you place the order as a means - but this is only a violation of the formula of humanity if, in so doing, you don't also treat them as an end in themselves.

'Act in such a way that you always treat humanity,
whether in your own person or the person of any
other, never simply as a means, but always at the same
time as an end.'

This style of deontological view might seem to give the correct account of our otherwise puzzling views of the trolley cases. It is wrong to kill David or to shove the fat man onto the tracks because those actions involve using them as a mere means to save the lives of others.

By contrast, in the case where you turn the trolley to save the five, you are not using the one person on the tracks as a means to save the five; they would be saved even without them there.

This style of deontological view is, in certain respects, much less demanding than consequentialist views. What would each theory say about the question of whether we are morally obliged to give to the poor?

This kind of view has a lot to recommend it. But it also faces serious objections; let's look at two.

The first is sometimes called the **paradox of deontology**. The view of ethics we are considering says that rights-violations are the core of morality; they are what we are morally obliged to avoid.

So now consider a situation in which I know that Dr. Evil is planning to use two innocent people as means to his ends. I can prevent this by lying to Dr. Evil, thereby using him as a means to my end of saving the innocents. Surely this is ok; rights-violations are supposed to be the bad thing, and I have prevented two while causing only one!

The deontologist rejects this kind of consequentialist thinking. You must not treat Dr. Evil as a mere means, even if by doing so you can prevent him from using others as a mere means.

The second is that there are plenty of cases where there seems to be a moral obligation to do or not do something, even though, either way, no one is treating anyone else as a mere means. Let's consider three examples of this.

The first are cases which involve no actions toward other human beings. Consider, for example, animal cruelty or environmental destruction. These actions involve treating no one as a mere means, but seem clearly wrong.

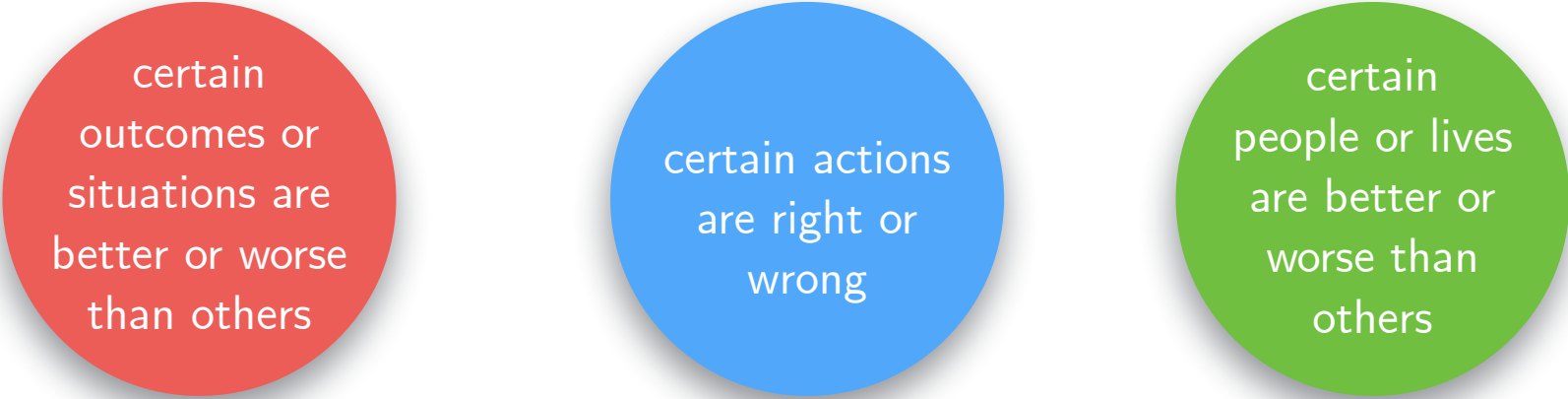
The second are cases of simple disregard. Suppose that I drive very quickly out of my neighborhood and run over a child. It does not seem that I used the child as a means to an end; but my action was surely wrong.

The third are cases in which there is a huge gap in the goodness of the outcome. Consider a version of the basic trolley case in which there are 1000 people on the track in front of one. Surely one is morally required to turn the trolley to kill the one. But you can let the trolley continue on its path without using anyone as a means to an end.

The third are cases in which there is a huge gap in the goodness of the outcome. Consider a version of the basic trolley case in which there are 1000 people on the track in front of one. Surely one is morally required to turn the trolley to kill the one. But you can let the trolley continue on its path without using anyone as a means to an end.

There is a troubling variant on this last kind of case. If we make the consequences bad enough, it can seem very plausible that some cases of using someone as a means to an end are morally permissible, and even required.

Take the case of the fat man. Suppose that by pushing him you can prevent the detonation of a hydrogen bomb which would kill millions. It seems plausible that in this case you are not just permitted, but required, to push him. But how can we get this result without opting for some form of consequentialism?



certain
outcomes or
situations are
better or worse
than others

certain actions
are right or
wrong

certain
people or lives
are better or
worse than
others

Our last approach takes as basic, neither the goodness or badness of situations nor the rightness or wrongness of actions, but the notion of a **good person** or **good human life**. This approach to our question is sometimes called 'virtue ethics.'

According to this kind of theory, if you are deciding between X and Y, you don't ask which would lead to the best consequences, and don't ask which would accord with the moral rules, but instead ask which would lead to you leading the best overall life.

The qualities which would contribute to your life being a good one are called **virtues**.

As with consequentialism and deontology, there are various different specific versions of virtue ethics. But as in those cases, I'll focus in on one, which is a version of a kind of theory called **perfectionism**. This kind of theory goes back to Aristotle.

As with consequentialism and deontology, there are various different specific versions of virtue ethics. But as in those cases, I'll focus in on one, which is a version of a kind of theory called **perfectionism**.

Let's forget about human lives for a second, and think about the lives of non-human animals. Consider the following two beasts:



We can ask, of either animal, what it would take for their life to go best. Do we get the same answer in the two cases?

We can ask, of either animal, what it would take for their life to go best. Do we get the same answer in the two cases?

It is natural to think that we do not. Lions are different kinds of things than dogs, and so the best kind of life for a lion will be very different than the best kind of life for a dog. In each case it depends on the nature of lions, and the nature of dogs.

Humans are different than both lions and dogs. So, Aristotle thought, to figure out what the best human life is, we have to ask: what is part of human nature?

When we ask this question, a number of plausible answers suggest themselves:

When we ask this question, a number of plausible answers suggest themselves:

Humans have bodies (of certain distinctive kinds).

Humans are social animals, who live with other humans.

Humans are strategic animals, who plan for the future.

Humans are creative animals, who make art & music.

Humans are rational animals, who try to get knowledge of the world

According to a perfectionist view, the best human life is the one which most perfects these aspects of human nature. The reason why these aspects matter is because they are part of what it is to be human — just as perfection in the hunting of gazelles matters for lions but not my dog, because this kind of hunting is in the nature of lions but not of my dog.

According to a perfectionist view, the best human life is the one which most perfects these aspects of human nature. The reason why these aspects matter is because they are part of what it is to be human — just as perfection in the hunting of gazelles matters for lions but not my dog, because this kind of hunting is in the nature of lions but not of my dog.

To each such aspect corresponds (at least) one virtue. For example, virtues to do with our social nature might include justice, generosity, patience, and kindness.

Does this mean that the best life for me is the same as the best life for you, even though we are different in any number of ways?

No. One of the ways in which people differ is in their capacity for perfection in various dimensions. You may have the capacity for great artistic creativity; I do not. This might mean that your greatest overall perfection would be achieved by devoting yourself to music, whereas mine would not be.

This leads lots of important questions unanswered. Sometimes trying to perfect one aspect of my nature conflicts with trying to perfect others. If I focused solely on trying to produce the best philosophy that I can, my physical well being would suffer and I would ignore the social aspect of my nature. Different perfectionists have different views about how these trade-offs should be managed.

Indeed, these unanswered questions are the source of the one of the central objections to virtue theory, which is that the theory does not seem to tell us as much as we would want a moral theory to tell us.

Consider the trolley problem. What does the virtue ethicist tell us to do in each version of the case? It is a little hard to say.

There are really two related problems here. The first is that virtue ethics does not issue the kind of clear guidance in genuine moral dilemmas that consequentialism and deontological theories seem to give us. If you are attracted to virtue ethics, you should try to think about a difficult moral decision, and see how thinking about the good human life and the virtues it involves might help.

The second is a worry that virtue theory will not give us the resources to see where our own moral intuitions go wrong. If our starting point is our view of the best human life, could virtue ethics ever show us that our assumptions about morality are fundamentally flawed?