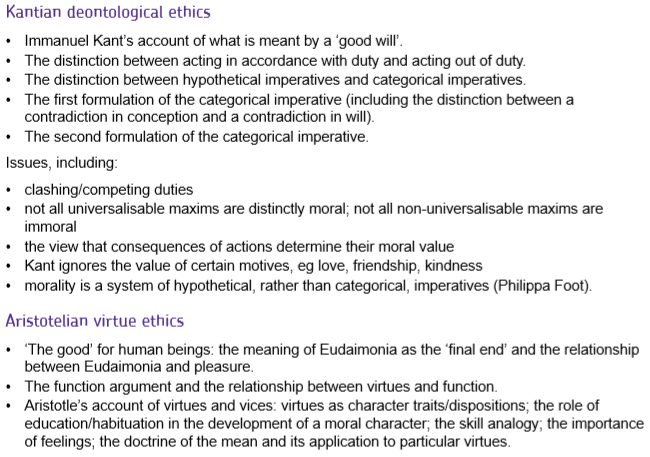
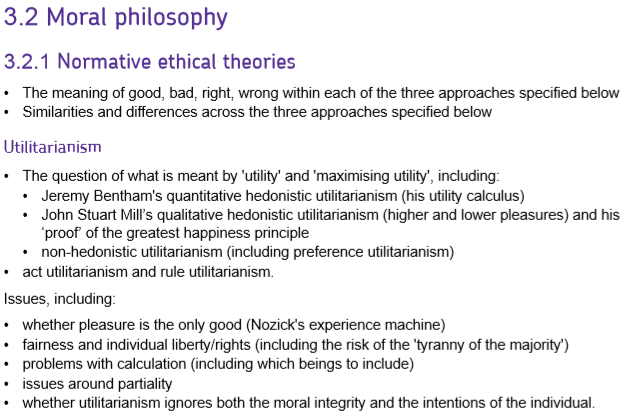
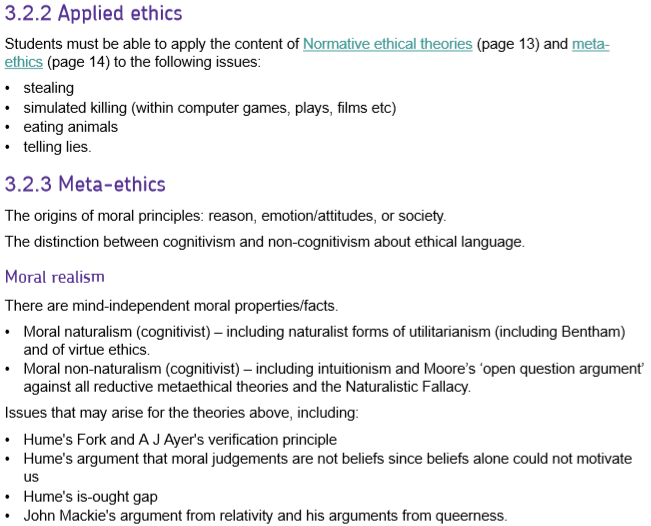
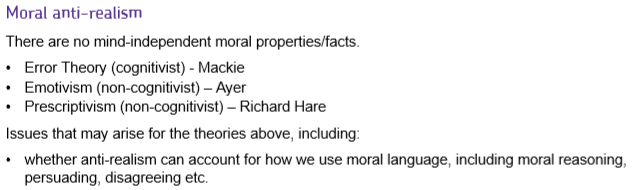


Moral Philosophy







1. Normative Ethical Theories

1.1. Utilitarianism

‘How should one live?’ It is not a trivial question, as Socrates says. Perhaps uniquely among animals, we human beings not only act, we also consider how we *should* act. We think that there are better and worse ways of acting, we reflect on our experience of making mistakes, and try to improve things. Much of this, of course, relates to our own self-interest – meeting our needs, successfully achieving our personal goals, and so on. But that is not all. We are social creatures, we live together, and our lives and actions affect the lives and actions of other people. How should we relate to one another, how should we treat one another? We are concerned not only for ourselves, but for other people as well, and how other people treat us is critical to our own happiness. How should each of us live so that each of our lives goes ‘best’? What is ‘good’ in life and how may we go about trying to attain it?

These questions form the basis for moral philosophy. Normative ethics is a branch of moral philosophy that aims to give us general guidance on what is morally right or wrong, what is good or bad. It develops theories about what people care about or what makes their lives go well, about how to live and what we should do.

# Hedonistic Act Utilitarianism

In its simplest form, utilitarianism is defined by three claims.

1. What is right? Actions are morally right or wrong depending on their consequences and nothing else. An act is right if it maximises what is good. This is ‘act consequentialism’.

2. What is good? The only thing that is good is happiness, understood as pleasure and the absence of pain. This is ‘hedonism’.

3. Who counts? No one’s happiness counts more than anyone else’s. This is a commitment to equality.

This is known as hedonistic act utilitarianism. If we put (1) and (2) together, we see that the theory claims that an action is right if it *maximises* happiness, i.e. if it leads to the greatest happiness of all those it affects. Otherwise, the action is wrong. Our actions are judged not ‘in themselves’, e.g. by what *type* of action they are (a lie, helping someone, etc.), but in terms of what *consequences* they have. Our actions are morally right if they bring about the greatest happiness.

‘Greatest happiness’ is comparative (great, greater, greatest). If an action leads to the greatest happiness of those it affects, no other action taken at that time could have led to greater happiness. So an action is right only if, out of all the actions you could have done, this action leads to more happiness than any other. Just causing *some* happiness, or more happiness than unhappiness, isn’t enough for an act to be morally right.

Act utilitarianism seems to provide a clear and simple way of making decisions: consider the consequences of the different actions you could perform and choose that action that brings about, or is likely to bring about, the greatest happiness. It makes complicated decisions simple and avoids appeals to controversial moral intuitions. The only thing that matters is happiness, and surely everyone wants to be happy. We can figure out empirically how much happiness actions cause, and so we can solve moral issues by empirical investigation.

# BENTHAM’S QUANTITATIVE HEDONISTIC UTILITARIANISM

Jeremy Bentham is considered the first act utilitarian. In *An Introduction to the Principles of Morals and Legislation*, he defended the ‘principle of utility’, also known as the ‘greatest happiness principle’. It is ‘that principle which approves or disapproves of every action whatsoever, according to the tendency which it appears to have to augment or diminish the happiness of the party whose interest is in question’. Or again, ‘that principle which states the greatest happiness of all those whose interest is in question, as being the right and proper … end of human action’. So Bentham claims that in judging actions to be morally right or wrong, we should take into account only the total amount of happiness that the action may produce. Likewise, in our own actions, we should aim to produce the greatest happiness we can.

## The meaning of ‘utility’

Utilitarianism is so-called because it is concerned with ‘utility’. Bentham explains what he means by ‘utility’, making the connection between utility and happiness:

By utility is meant that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness, (all this in the present case comes to the same thing) or (what comes again to the same thing) to prevent the happening of mischief, pain, evil, or unhappiness to the party whose interest is considered.

He then clarifies what he means by ‘interest’:

A thing is said to promote the interest, or to be *for* the interest, of an individual, when it tends to add to the sum total of his pleasures: or, what comes to the same thing, to diminish the sum total of his pains.

So, something has ‘utility’ if it contributes to your happiness, which is the same as what is in your interest. And happiness is pleasure and the absence of pain. The claim that pleasure, as happiness, is the only good is known as hedonism. Bentham goes on to list 14 ‘families’ of pleasure, such as sensory pleasure, the pleasures of exercising one’s skills, the pleasures of having power, the pleasures of memory, and the pleasures of benevolence. He also lists 12 families of pain, many deriving from similar sources as pleasure.

## ‘Measuring Pleasure and Pain’

Bentham argued that we can measure pleasures and pains and add them up on a single scale by a process he called the ‘felicific calculus’ (‘felicity’ means happiness), also known as the ‘utility calculus’. If a pleasure is more intense, will last longer, is more certain to occur, will happen sooner rather than later, or will produce in turn many other pleasures and few pains, it counts for more. In thinking what to do, you also need to take into account how many people will be affected (the more you affect positively, and the fewer you affect negatively, the better). The total amount of happiness produced is the sum total of everyone’s pleasures produced minus the sum total of everyone’s pains. As this demonstrates, Bentham took a *quantitative* approach to happiness.

## Discussion

The reasons to believe utilitarianism rest in its intuitive appeal. Everyone cares about happiness (Bentham claims that the *only* things that motivate people are pleasure and pain). Morality is about how to act, so it better be about what motivates us. So it is about happiness. If happiness is good, then surely it is reasonable to think that more happiness is better. So we should maximise happiness. And until we have a good reason to think otherwise, treating people as equal is an appealing moral starting point.

However, Bentham’s normative ethical theory may also strike us as too simple. For example, is happiness the *only* thing that matters morally, i.e. is hedonism the correct theory of what is good? Should we really bring about the greatest happiness in all situations, even when we have to violate someone’s rights to do so? Or again, ‘more’ happiness might be better than ‘less’, but can we really ‘add up’ how much happiness an action will cause? So even if we think that morality has *something* to do with happiness, we might want to reject utilitarianism.

# Mill on Utilitarianism

John Stuart Mill begins his book *Utilitarianism* by remarking on how surprising it is that, with all the developments in knowledge over the last two millennia, there is still little agreement on the criterion for right and wrong. In science, we start from particular observations and work out the laws of nature from them. But our usual method of empirical induction doesn’t work in ethics. Part of the difficulty is that we can’t easily infer the principles of morality from particular cases, because we first need to know the principles in order to judge whether an action is right or wrong.

However, we shouldn’t exaggerate the disagreement. Many philosophers agree that morality involves moral laws and they agree on what many of these laws are (e.g. concerning murder, theft, harming others, betrayal, etc.), even if they disagree about why these are moral laws. Mill then remarks that people’s moral approval and disapproval is, as a matter of fact, strongly influenced by the effects of actions on their happiness. So the principle of utility has played a significant role in forming moral beliefs, even if this hasn’t been recognised.

When Mill wrote *Utilitarianism*, there was a lot of misunderstanding of what utilitarians actually believed. Mill clarifies what utilitarianism is – what it really claims – by considering and replying to 11 objections which he argues arise from misunderstanding the theory.

1. ‘Utility’ means what is useful, not what is pleasurable. Utilitarianism therefore ignores the value of pleasure.

Reply: Obviously a misunderstanding. Mill reasserts Bentham’s central claims. First, ‘actions are right in proportion as they tend to promote happiness, wrong as they tend to produce the reverse of happiness’. Second, ‘By happiness is intended pleasure, and the absence of pain; by unhappiness, pain, and the privation of pleasure.’ Third, ‘pleasure, and freedom from pain, are the only things desirable as ends’. Like Bentham, then, Mill accepts and defends hedonism.

1. We do not need happiness and many wise and noble people have lived without it.

Reply: True, but what have noble people sacrificed their happiness for? Surely, it is the happiness of others. If not, then what a wasted sacrifice! Utilitarianism recognises the virtue of sacrificing your happiness for others – the aim remains to increase the total happiness in the world.

1. Utilitarians make right and wrong depend upon the agent’s happiness.

Reply: Another obvious misunderstanding. It is the happiness of everyone that is the criterion of right action. For this reason, we should organise society and raise children in such a way that each person feels that their own happiness is bound up with the happiness of others, that they are made happy by making others happy.

1. Utilitarianism is a godless theory.

Reply: It isn’t. Utilitarianism can easily be made compatible with Christian teachings about God. (Given his social context, Mill mentions only Christianity.)

1. Utilitarianism will lead to people sacrificing moral principles for ‘expedient’ immoral action.

Reply: ‘Expedient’ usually means either what is in the person’s own interest or in the short-term interest, as when someone lies to get out of a tricky situation. Where an action is ‘expedient’ in this sense and sacrifices the greater happiness of people generally, then utilitarianism condemns it.

# HIGHER AND LOWER PLEASURES

Mill argues that the claim that utilitarianism degrades human beings misunderstands what human beings take pleasure in. Some types of pleasure are ‘higher’ than others, more valuable, more important to human happiness, given the types of creatures we are and what we are capable of.

Which pleasures? How can we tell if a type of pleasure is more valuable (quality) than another, rather than just more pleasurable (quantity)? The answer has to be to ask people who know what they are talking about. If everyone (or almost everyone) who has experience of two ***types*** of pleasure prefers one type to the other, then the type that they prefer is more valuable. To ensure that they are considering the quality and not quantity of the pleasure, we should add another condition. A pleasure is higher only if people who have experience of both types of pleasure prefer one even if having that pleasure brings more pain with it, or again, even if they would choose it over a greater quantity of the other type of pleasure.

Mill argues that, as long as our physical needs are met, people will prefer the pleasures of thought, feeling and imagination to pleasures of the body and the senses, even though our ‘higher’ capacities also mean we can experience terrible pain, boredom and dissatisfaction. For example, as Tennyson said in his poem ‘In Memoriam A.H.H.’, ‘’Tis better to have loved and lost than never to have loved at all’. We can say the same about intelligence and artistic creativity – better to have the pleasures that they bring, even though they cause us pain and distress, than to be unintelligent or lack creativity.

Thus Mill compares a human being with a pig. As human beings, we are able to experience pleasures of deep personal relationships, art and creative thought that pigs are not. We can experience new and deeper kinds of pain as a result. Yet we don’t think that the possibility of pain would be a good reason for choosing to be a well-looked-after pig, rather than a human being. ‘It is better to be a human being dissatisfied than a pig satisfied’. This preference, Mill thinks, derives from our sense of dignity, which is an essential part of our happiness.

In introducing this distinction between higher and lower pleasures, Mill rejects Bentham’s felicific calculus – the idea that we can simply ‘add up’ how much happiness some action produces, and adds the element of quality to the quantitative analysis of happiness that Bentham puts forward.

## Discussion

It is important to note that if Mill’s prediction here is wrong – if people with the relevant experience do not prefer the pleasures of thought, feeling and imagination to other pleasures, then these are not higher pleasures. So we can object that people do *not* reliably pursue the ‘higher’ pleasures of thought, feeling and imagination instead of the ‘lower’ pleasures related to the body and the senses.

Mill accepts the point but argues that it is no objection. First, there is a difference between preference and action. We can choose what we know to be less good, whether from weakness of will or laziness or other factors. We still recognise that what we did not choose is more valuable.

Second, appreciating the higher pleasures can be more demanding. Our ability to experience higher pleasures can be undermined by hard work, lack of time, infrequent opportunities to experience them, and so on. We may seek lower pleasures simply because those are more readily available to us.

Not just anyone’s preference counts as deciding whether a pleasure is ‘higher’ or ‘lower’. As with any question, we need to consult people who know what they are talking about. Having been to an art gallery once doesn’t count as having experienced the pleasures of art, and listening to just one pop song doesn’t count as having experienced the pleasures of pop music. Mill says that one pleasure is higher than another if almost everyone who is ‘competently acquainted’ with both prefers one over the other.

# Is pleasure the only good?

## Smart on hedonistic and non-hedonistic utilitarianism

As we have seen in the case of Bentham and Mill, utilitarians can disagree about what is good, and so about which consequences or ‘states of affairs’ are best. Bentham thinks that we should concern ourselves only with the quantity of happiness caused while Mill thinks matters are more complicated, and we need to take into account the quality of happiness as well (higher v. lower pleasures). This question of what is good is the focus of J.J.C. Smart’s discussion in his book *Utilitarianism*. Is only pleasure good? Is only happiness, understood as involving both quantity and quality of pleasure, good? Is anything other than pleasure or happiness good? In connection with the last option, Smart mentions G E Moore, who thought that some other states of mind, aside from pleasure, are also good, such as knowledge. Smart calls Moore an ‘ideal’ utilitarian, as opposed to the ‘hedonist’ utilitarianism of Bentham and Mill.

How might Bentham respond to Mill’s distinction between higher and lower pleasures? He could say that our preference for higher pleasures is not because such pleasures are qualitatively better, but because they are more ‘fecund’ – they are likely to produce *more* pleasure either for others or for ourselves in the long term. For example, if we enjoy thought, then we may produce greater benefits for society, e.g. through scientific research.

Yet, Smart claims, Mill seems right that pleasures are not all equal – if they were, a contented sheep is as good as a contented human, so perhaps, to maximise happiness, we should decrease the human population and increase the number of cared-for sheep! It seems that most of us, at least, do prefer the pleasures of thought, feeling and imagination – perhaps because human beings are intellectual creatures: we owe our survival in evolutionary terms to our intellect, not our bodily strength or speed.

To try to understand our preferences for certain pleasures further, Smart asks us to imagine a scenario in which someone is wired up to electrodes in his brain, so that just by pressing a button, he could give himself intense sensual pleasure reliably and with no ill-effects. What should we say if he came to prefer this to anything else, and spent all his spare time doing this? What if we knew that most people would feel the same if they were wired up? Is this a picture of a good or happy life, one we should bring about as the maximisation of happiness?

Smart argues that while we can understand that the electrode operator is *contented*, and even that we would be contented if we became electrode operators, we simply don’t want to become electrode operators. We want to do more with our lives than that, to achieve things, and if someone were to force us to become electrode operators, we would not now thank them for it, but be made unhappy by the prospect. The same applies to Mill’s dictum ‘Better to be Socrates dissatisfied than a fool satisfied’. Socrates would not want to be a fool, even knowing that after he became a fool, he would be contented.

Smart argues that our responses here are not a matter of the quality of the pleasure, as Mill thinks, but our *attitude* towards it. To say that someone is *happy*, and not just contented, is to express approval of their pleasures. Because we think a life as an electrode operator is wasted, we do not think such a person ‘happy’, only contented, and we would not want such a life for ourselves. Of course, Mill rejects this analysis: he says that when we judge pleasures as higher or lower, we should put aside all considerations of morality, of whether we *ought* to prefer the pleasure to another. Smart claims we don’t do this – we prefer certain pleasures to others and our approval of them is an aspect of this.

Smart concludes that Mill’s position is not a form of *pure* hedonism after all. Pure hedonism recognises pleasure as the only good and sole criterion for what we ought to bring about. Such a view can only defend the place of higher pleasures in terms of their fecundity; they are not ‘higher’ at all. But Mill’s concept of happiness is evaluative, because happiness is pleasure of which we approve. And so it is not pure hedonism.

## Is pleasure always good?

Mill and Bentham agree that pleasure is always good. But imagine a world in which there is just one person, who believes (falsely) that elsewhere in the world, other people are being tortured. This thought delights him a great deal; he is a sadist. Is his pleasure bad? Would the world be a morally better place if this belief caused him sadness?

Smart defends the hedonist position: it is good that he feels pleasure at the thought of their suffering, because at least he is happy, and no one else is, in fact, suffering. Our difficulty in accepting this is that in the real world, sadistic thoughts and pleasures often cause *actual* suffering. In other words, they lead to morally bad consequences. But we need to separate this from whether they are bad in and of themselves, ‘intrinsically’. No pleasure, Smart claims, is intrinsically bad; it is only ever bad if it causes pain (to the person themselves or others).

# Nozick’s experience machine

In his discussion of the electrode machine, Smart contrasts our desire for pleasure with our desire for achievements, and then talks of happiness as not only involving contentment, but approving of our contentment. We can question whether this really gets to the bottom of our desire for achievement.

In *Anarchy, State and Utopia*, Robert Nozick asks us to imagine being faced with the chance of plugging in to a virtual reality machine. This machine will produce the experience of a very happy life, not only with many and various pleasures and few pains, but (the experience of) many successful achievements. If we plug in, we will not know that we are in a virtual reality machine. We will believe that what we experience is reality. However, we must agree to plug in for life or not at all.

Nozick argues that most of us would *not* plug in. We value being in contact with reality, even if that makes us less happy, even if we experience fewer achievements. But we can’t understand this in terms of the ‘pleasure’ of being in touch with reality, or a preference for certain types of pleasure over others, because if we were in the machine, we would still experience this pleasure (we would believe we were in touch with reality). Instead, what we want is not a *psychological state* at all; it is a *relation* to something outside our minds. Smart is right that we want achievements; but we want *real* achievements, not just the psychological state of experiencing an achievement. Nozick concludes that we cannot understand what is good just in terms of our subjective psychological states, such as pleasure.

# Preference utilitarianism

One solution to the complications around pleasure and happiness facing hedonistic utilitarianism is offered by preference utilitarianism. Preference utilitarianism is a form of non-hedonistic utilitarianism that argues that what we should maximise is not pleasure, but the *satisfaction of people’s preferences (desires)*.

1. If Nozick is right, we prefer to be in touch with reality, but not because it brings us pleasure. Having this preference satisfied is valuable. For a preference to be satisfied, it must be satisfied in reality. It is not enough that the person thinks their preference is satisfied.
2. We can also appeal to preferences to explain Mill’s claims about higher and lower pleasures. He defends the distinction in terms of what people prefer. However, rather than talk about the value or quality of types of pleasure, we could argue that whatever people prefer is of more value to them – whether or not most people would prefer pleasures related to thought, feeling and imagination.
3. The satisfaction of many of our preferences will bring us pleasure, but many will not. For instance, Bentham and Mill do not distinguish between producing happiness and decreasing pain. But are these morally equivalent? If people more strongly prefer not to suffer pain than to be brought pleasure, then that would explain the thought that it is more important not to cause harm. Or again, we can also argue that people have preferences about what happens after their death, e.g. to their possessions, and it is important to satisfy these as well, even though this cannot bring them any pleasure.

In sum, preference utilitarians can argue that they offer a more unified account of what is valuable than hedonist utilitarianism. We can continue to claim that happiness is the only good if we now understand happiness as one’s desires being satisfied. Pleasure is important, when it is, because it results from satisfying people’s preferences.

They can also argue that they simplify Smart’s theory. Smart argues that happiness is good, but he understands happiness in terms of those pleasures we approve of. But Smart offers no further explanation of this approval. People simply do or don’t approve of certain pleasures. But shouldn’t we approve of what is good? In other words, shouldn’t a theory of what is good say what we *should* approve of, rather than explain what is good in terms of what we *do* approve of (with no further reason)? Preference utilitarianism can say that what is good is maximising the satisfaction of people’s preferences. That many people prefer certain pleasures to others is enough reason to prioritise those pleasures; approval isn’t relevant.

Mill’s ‘proof’ of utilitarianism

Utilitarianism claims that the only thing that is good, and so the only thing that our actions and lives should aim at, is happiness. Why believe this, and how can we know what is good? In *Utilitarianism*, John Stuart Mill seeks to prove that utilitarianism is right about happiness being the only good. The proof has two stages. In the first stage, Mill argues that happiness is good. In the second stage, he argues that it is the only thing that is good.

# Stage 1: happiness is good

Mill argues that you can’t strictly ‘prove’ that something is good or not. That is, it is not something that you can deduce from other premises. This is normal for ‘first principles’ in any area of knowledge, and a claim about what is ultimately good is a first principle in ethics. Nonetheless, we can give a reasoned argument about what is good.

First, some terminology. What is good is what we should aim at in our actions and lives. So what is good is an ‘end’ – the purpose – of our actions. Philosophers understand actions in terms of means and ends. Ends are why you do what you do; means are how you get it. So I might cross the street to post a letter. My end is posting the letter, my means is crossing the street. Now, of course, my posting a letter is also a means to an end, the end of communicating with someone. This, too, may be a means to an end. Perhaps I am asking them for a favour. So I cross the street in order to post the letter in order to ask someone a favour. What is the end of asking them for a favour? What am I ultimately aiming at?

What we should aim at is what is desirable, says Mill. So what he wants to show is, first, that happiness is desirable, and second, that only happiness is desirable. If he is right, then the answer to our previous question about why I am asking for a favour will be ‘happiness’.

Since we can’t deduce what is good, we have to appeal to evidence. Mill’s argument that happiness is good has three parts.

1. ‘The only proof capable of being given that an object is visible is that people actually see it … In like manner … the sole evidence … that something is desirable is that people do actually desire it ….’
2. ‘No reason can be given why the general happiness is desirable, except that each person … desires his own happiness.’
3. ‘This, however, being a fact, we have not only all the proof the case admits of, but all which it is possible to require, that happiness is a good: that each person’s happiness is a good to that person, and the general happiness, therefore, a good to the aggregate of all persons ….’ Put more clearly, each person takes their own happiness to be good, and so, adding each person’s happiness to that of others, the happiness of everyone – the general happiness – is good for people in general.

## Clarifying the argument

G. E. Moore objected that Mill commits the fallacy of equivocation in this argument, confusing two meanings of a word. The word ‘desirable’ has two meanings. Its usual meaning is ‘worthy of being desired’. Anything desirable in this sense is good. This is the sense it has in (2), since Mill is arguing that the general happiness is good. But another meaning could be ‘capable of being desired’. To discover what is capable of being desired, look at what people desire. This is the sense it has in (1), it seems, since Mill links what is desirable to what people desire. But what people actually desire is not the same as what is worthy of being desired (good). People want all sorts of rubbish! Mill has assumed that what people desire just is what is good; he hasn’t spotted that these are distinct meanings of ‘desirable’.

But Moore’s objection misinterprets Mill’s argument. Mill is asking ‘What evidence is there for thinking that something is worthy of being desired?’ He argues that people in general desire happiness. Unless people in general desire what is not worth desiring, this looks like good evidence that happiness is desirable. Is there anything that everyone wants that is not worth wanting? If we look at what people agree upon in what they desire, we will find evidence of what is worth desiring. Everyone wants happiness, so it is reasonable to infer that happiness is desirable (good).

Other philosophers have objected that Mill commits the fallacy of composition in (3). This is a fallacy of inferring that because some part has a property, the whole of which it is a part also has that property. Mill seems to be saying that because each person desires their own happiness, everybody desires everybody’s happiness (the general happiness). But this doesn’t follow. For example, suppose that every girl loves a sailor (substitute ‘own happiness’). From the fact that for each girl, there is some sailor that she loves, we cannot infer that there is one sailor (substitute ‘general happiness’) which every girl loves.

But this is also a misinterpretation of Mill’s argument. At no point does Mill feel that he needs to defend the idea of impartiality in ethics. He simply assumes that ethics is concerned with what is good in general. He is not trying to infer that we ought to be concerned for others’ happiness. Having argued that happiness is good, it follows from his assumption that ethics is impartial that we should be concerned with the general happiness.

# Stage 2: only happiness is good

The claim that happiness is good is relatively uncontroversial. It is much more controversial to claim that it is the only good. Mill must argue that everything of value – truth, beauty, freedom, etc. – derives its value from happiness.

Now if people only ever desired happiness, he could use the previous argument to show that happiness is the only good. But clearly, people desire many different things. Of course, we may desire many things as a means to happiness, such as buying a nice house or having a good job. But it isn’t obvious that everything we desire is a means to happiness. For example, we want truth (being in touch with reality) and not because it has some psychological effect on us. So going by the evidence, many different things, and not only happiness, are good.

Mill’s response is to clarify further what happiness is. Happiness has many ‘ingredients’, such as truth and freedom, and each ingredient is desirable in itself. We can explain this in terms of a distinction between ‘external means’ and ‘constitutive means’ to an end. We usually think of the relation between means and end as an instrumental relation; i.e. that performing the means achieves the further, independent end. Think about having a good holiday. Suppose you have to get up very early in order to catch the plane. You do this in order to have a good holiday, but it isn’t part of having a good holiday. Getting up early is an external means to the end. But there is also another relation between means and ends, a constitutive relation. Later on, you are lying on the beach in the sun, listening to your favourite music. Are you doing this ‘in order’ to have a good holiday? Not in the same sense. This just is having a good holiday at the moment. Lying on the beach is a constitutive means to the end of having a good holiday. Having a good holiday is not something ‘further’ or additional that you achieve by lying on the beach. In these circumstances, here and now, it is what ‘having a good holiday’ amounts to.

The same applies to happiness, Mill argues. For example, when someone desires to know the truth ‘for its own sake’, their knowing the truth doesn’t cause their happiness as some further and separate thing. Rather, in this situation, their happiness consists in their knowing the truth. Knowing the truth for its own sake is part of happiness for them. So, Mill claims, whatever we desire for its own sake is part of what happiness is, for us.

Why believe this? Mill argues that to desire something just is to find it pleasant. It is, he says, ‘physically and metaphysically impossible’ to desire something that you don’t think is a pleasure. As pleasure is happiness, we only desire happiness, and happiness is the only good.

But is Mill right about the relation between pleasure and happiness? Preference utilitarianism allows that, if we understand pleasure as a psychological state (which we should), we can desire something without desiring it ‘as pleasant’. What we desire is part of our happiness because happiness is the satisfaction of desires, not because happiness is pleasure which is caused by the satisfaction of desires. J.J.C. Smart adds an evaluative element to happiness. For example, we are made happy by knowing the truth because we approve of knowing the truth, and not simply because it brings us pleasure (in the sense of contentment).

Rule utilitarianism

Act utilitarianism judges the rightness or wrongness of an act by its consequences. It claims:

1. Actions are morally right or wrong depending on their consequences and nothing else. An act is right if it maximises what is good.
2. The only thing that is good is happiness.
3. No one’s happiness counts more than anyone else’s.

If we put (1) and (2) together, we see that the theory claims that an action is right if it *maximises* happiness, i.e. if it leads to the greatest happiness of all those it affects. Otherwise, the action is wrong.

Rule utilitarianism agrees with claims (2) and (3), but disagrees with (1). It claims that an action is right if, and only if, it complies with those rules which, if everybody followed them, would lead to the greatest happiness (compared to any other set of rules). In other words, rule utilitarians do not look at the consequences of individual actions, but at the consequences of people following rules. A rule is morally right if following it leads to greater happiness than following an alternative rule. An action is morally right if it complies with a rule that is morally right.

# Smart on rule utilitarianism

In *Utilitarianism*, J.J.C. Smart objects that rule utilitarianism amounts to ‘rule worship’. The point of the rules is to bring about the greatest happiness. If there is a situation in which breaking the rule will lead to more happiness than following the rule, what reason could we have for following the rule? If I know e.g. that lying in a particular situation will produce more happiness than telling the truth, it seems pointless to tell the truth, causing unhappiness. The whole point of the rule was to bring about happiness, so there should be an exception to the rule in this case. The alternative to ‘everyone obeys the rule’ is not ‘no one obeys the rule’. Clearly, the best course of action is for some people to break the rule sometimes, in situations in which breaking the rule leads to more happiness than following the rule.

Rule utilitarians could respond by saying that we should amend the rule to allow the exception in such cases. For example, ‘Don’t lie’ should become ‘Don’t lie unless telling the truth will hurt someone’s feelings’. However, life is complicated. Whenever a particular action causes more happiness by breaking the rule than by following it, we should do that action. Smart points out that, if we keep amending the rules like this, there is no difference between what an act utilitarian would recommend and what a rule utilitarian would recommend. Furthermore, and more seriously for rule utilitarianism, if we try to add all the possible amendments to the rules in order to make acting on them always produce the greatest happiness, we will end up with just one rule, namely to maximise the greatest happiness. Because there are so many situations in which breaking a rule might lead to greater happiness than following it, no other rule can be certain of maximising happiness. And so rule utilitarianism collapses into act utilitarianism. It isn’t a distinct theory at all.

# Rule utilitarianism developed

To answer this objection, rule utilitarians can do two things. First, they could provide some reason for following rules even when doing so doesn’t maximise happiness. Second, they could argue that despite recommending that we follow rules when doing so doesn’t maximise happiness, it is still a better theory than act utilitarianism. In fact, rule utilitarians have provided arguments that support both points at once.

Their first argument is that morality should be understood as a set of rules. It needs to provide general guidance over the long term, a way of thinking that people can remember and rely on. So rather than considering actions individually, we need to take the bigger picture in thinking about right and wrong. This is just a reflection on what morality must be, psychologically speaking. Since we are to have rules, then the aim of these rules should be to maximise happiness. And so actions are right when they follow a rule that maximises happiness overall – even when the action itself doesn’t maximise happiness in this particular situation.

Second, rule utilitarians argue that their theory has a number of advantages over act utilitarianism.

1. Problem with calculation: Act utilitarianism requires that we work out the consequences of an action for human happiness. But is this realistic? How can we know or work out the consequences of an action, to discover whether it maximises happiness or not? Surely this will be too difficult and too time-consuming for us to do.

But according to rule utilitarianism, we don’t have to work out the consequences of each act in turn to see if it is right. We need to work out which rules create the greatest happiness, but we only need to do this once, and we can do it together. (This is what Mill says human beings have done over time, giving us our customary moral rules.) Rule utilitarianism gives rules a formal place in its theory of whether an action is right.

1. Fairness, liberty and rights: Act utilitarianism does not rule out any type of action as immoral. There are no constraints on pursuing the greatest happiness. For example, if torturing a child produces the greatest happiness, then it is right to torture a child. Suppose a group of child abusers only find and torture abandoned children. Only the child suffers pain (no one else knows about their activities). But they derive a great deal of happiness. So more happiness is produced by torturing the child than not, so it is morally right. This is clearly the wrong answer. More generally, act utilitarianism, we may object, doesn’t respect individual rights or liberty, because it doesn’t recognise any restrictions on actions that create the greatest happiness.

By contrast, rule utilitarianism can plausibly argue that a rule forbidding torture of children will clearly cause more happiness if everyone followed it than a rule allowing torture of children. So it is wrong to torture children. More generally, individuals have rights, which are rules, because if people have to follow these rules (respect people’s rights), that leads to the greatest happiness. Rules requiring fairness and justice will produce greater happiness in the long run than rules that do not.

1. Partiality: Many of the things that we do to make people happy are aimed at specific other people, e.g. our family and friends. But act utilitarianism argues that in our decisions, we need to consider the greatest happiness that our actions could create, and this requires us to consider the happiness of each person equally. In other words, we should be impartial. If act utilitarianism is right, it seems we should spend much less time with the particular people we love and more time helping people who need help.

However, rule utilitarians can argue that a rule that allows partiality to ourselves, our family and friends will create more happiness than a rule that requires us to be impartial all the time. This explains the moral importance of partial relationships – they are necessary to happiness. Of course, we shouldn’t be completely partial. We still need to consider the general happiness, but we only need to act in such a way that, if everyone acted like that, would promote the greatest happiness. For example, in the case of charity, I only need to give as much to charity as would be a ‘fair share’ of the amount needed to really help other people. This combination of partiality and impartiality respects both our natural inclinations and the demands of morality.

1. Moral integrity: Having integrity involves acting according to your own values, sticking to them especially in the face of temptation or other situations that would make it easier to do something you consider wrong. Just as act utilitarianism appears to require us to set aside our partiality, it can also seem to require that we set our moral values in order to maximise happiness. In other words, utilitarianism attacks our moral integrity.

The rule utilitarian may provide a similar argument for integrity as partiality. The best rule, perhaps, will be one that allows exceptions to other rules, i.e. a rule that allows people not to follow other rules if they conflict with one’s integrity.

1. Intentions: Act utilitarianism claims that an action is right if it leads to the greatest happiness. It does not, therefore, recognise the moral value of our intentions in acting as we do. But whether someone intends to harm us or not – whether or not they do harm us – makes a big difference to how we respond to their action.

Rule utilitarianism can explain the importance of intentions by thinking about what it is for an action to comply with a rule. It is not just a matter of what its consequences are. For example, a person kills someone else. But was this murder or manslaughter or something else again? If the person intended to kill someone, and that is what they wanted to bring about, that is very different than if the killing was accidental or if the person was only intending to defend themselves against an attack. A rule prohibiting murder is not a rule prohibiting self-defence, even if self-defence may lead to death. So whether an action complies with a rule or not depends, at least in part, on the individual’s intention in performing the act. Two actions may have the same consequences, and yet one is right and one is wrong, because they are different types of action and fall under different moral rules, because the person’s intention in each was different. And so rule utilitarianism can recognise that the thought does count.

These developments in the theory enable the rule utilitarian to respond to Smart’s objection. If people try to follow act utilitarianism, this will lead to less happiness in the long term. For instance, people will make mistakes in working out what the consequences of their action for happiness will be. They will no longer feel secure in their rights and there will be pressure on personal relationships to be given up in favour of impartiality. And we will fail to make important distinctions between different types of action. So we may have to give up some happiness here and now to be more certain of the greatest happiness overall. Following rules (in fact, just the existence of the rule) provides this certainty.

All this provides reasons to have and follow rules even when doing so doesn’t lead to the greatest happiness on occasion. An act is not made right by maximising happiness but by being in accordance in a rule. If there is a conflict between following the rule and maximising happiness (e.g. telling a lie when ‘no harm is done’), there should be no temptation for the utilitarian to say that we should break the rule. Even if the act does maximise happiness, this doesn’t make it right.

# Objections

We may question whether the replies that rule utilitarianism gives above are persuasive.

First, we can object that rule utilitarianism still fails to understand what is important about partiality. Rule utilitarianism permits partiality and it encourages it insofar as partiality contributes to happiness. But partiality is good, according to rule utilitarianism, because it contributes to the general happiness. The importance of the individual, and making someone happy just for their sake, still seems to be missing. For instance, if I form friendships because this maximises happiness, does that respect and value my friends for themselves, as the particular people they are? Consider this example from Bernard Williams. Suppose a man is in a boating accident with both his wife and a stranger. Neither can swim, and he can only rescue one. We might think that he should simply rescue his wife. But if he thinks, ‘Rescuing my wife is in accordance with a rule that will lead to greater happiness than a rule that requires me to rescue the stranger’, this seems to miss the particular importance that being married has, including its moral importance.

Second, a rule that protects integrity could prove very problematic. If what someone values goes strongly against promoting the greatest happiness, then morality itself may seem in conflict with their integrity. Is it possible, for instance, that someone finds meaning in their life through making as much money as possible, without constraints? The rule utilitarian must insist that people’s values, and so their integrity, are broadly in line with what the set of rules that maximises happiness generally requires. Put another way, morality itself probably needs to be part of people’s commitments for rule utilitarianism to respect their integrity. But now, is this an objection any more? Or simply a recognition of the legitimate demands that morality makes upon us?

Rule utilitarianism also faces an objection based on its theory of the good. It agrees with act utilitarianism that happiness is the only good. But is this right? Suppose morality is a set of rules. Should that set of rules aim to maximise happiness, or are there other important values, such as telling the truth, that matter independent of the pleasure they bring (or preferences they satisfy)?

A new objection that rule utilitarianism faces, which act utilitarianism doesn’t, is whether all of morality can be summed up by rules. Isn’t life too complicated for this? If so, we will need a different theory to explain what the right thing to do is when there are no rules that apply. The rule utilitarian has a simple answer to this objection, though we can question whether it is adequate. One of the rules is ‘When no other rules apply, do that action that maximises happiness’.

Objection 1 - Calculation

Act utilitarianism seems to offer a clear and straightforward way of discovering what is right and wrong. We need to consider how much happiness an action will cause. But is it possible to work out the consequences of an action for human happiness? How can we know or work out the consequences of an action, to discover whether it maximises happiness or not? Surely this will be too difficult and too time-consuming for us to do. Bentham’s felicific calculus says we should consider for each possible pleasure produced by each possible action is more intense, will last longer, is more certain to occur, will happen sooner rather than later, or will produce in turn many other pleasures and few pains, than the pleasures produced by other actions. We also need to take into account how many people will be affected. All this is, in practice, mind-boggling, and we just can’t get the relevant information (how intense each affected person’s pleasure or pain will be, how long it will last, what other pleasures or pains it might cause in turn, etc.).

Preference utilitarianism might try to claim an advantage here. It is easier to know whether someone’s preference has been satisfied than how much pleasure someone experiences. But this is very little improvement if we still need to compare the strength of different people’s preferences, whether satisfying one preference leads to further preferences being satisfied, and so on.

However, the objection misrepresents what utilitarians say. Bentham does not say that an action is right if it actually maximises happiness. He says it is right according to ‘the tendency which it appears to have’ to maximise happiness. We don’t need to be able to work out the consequences precisely. An action is right if we can reasonably expect that it will maximise happiness. He also says that the felicific calculus need not be ‘strictly pursued’ before each decision or moral judgement. It just needs to be ‘kept in view’.

This still means we must be able to work things out roughly. Mill thought this was still too demanding. Happiness is ‘much too complex and indefinite’ a standard to apply directly to actions. But we don’t need to try, he claims, because over time, people have automatically, through trial and error, worked out which actions tend to produce happiness. This is what our inherited moral rules actually are: ‘tell the truth’, ‘don’t steal’, and ‘keep your promises’ are embodiments of the wisdom of humanity that lying, theft and false promising tend to lead to unhappiness.

Mill calls these moral rules ‘secondary principles’. It is only in cases of conflict between secondary principles (e.g. if by telling the truth you break your promise) that we need to apply the greatest happiness principle directly to an action. We shouldn’t attempt to calculate happiness directly unless we have such a conflict. Only in cases of conflict will there be genuine exceptions to these rules.

Of course, our inherited morality still makes mistakes in what it thinks will or won’t contribute to general happiness. So we can improve on the rules that we have. But saying this is quite different from saying that we have to consider each action from scratch, as though we had no prior moral knowledge.

# Which beings’ happiness should we include?

A number of these issues about calculation, and whether Mill’s responses really solve the problem, come into sharp relief when we consider which beings which should be included when calculating the happiness or unhappiness caused by an action.

Bentham was aware that his identification of happiness as the only good had some radical implications. In *An Introduction to the Principles of Morals and Legislation*, He argued that the question about who or what to consider when looking at the consequences of our actions is not ‘Can they reason? nor, Can they talk? but, Can they suffer?’

Utilitarianism says that happiness is good, not just that the happiness of humans is good. If happiness is good, then it is good no matter what creature feels it. There is nothing in the theory that gives us a reason to privilege human happiness over the happiness of non-human animals. So it seems that the logic of utilitarianism requires us to take as much account of beings that are not human as of human beings. The third condition of act utilitarianism – that the happiness of each matters equally – does not stop at the boundary between human and non-human.

This line of thought has been more recently developed by Peter Singer in his book *Animal Liberation*. We do not think that it is right to treat women worse than men just because they are women (this is sexism), nor to treat one race worse than another (this is racism). Likewise, it is wrong to treat animals as unequal just because they are not human. This is ‘speciesism’.

We can object that with women and men, and different races, there is no difference in those important capacities – reason, the use of language, the depth of our emotional experience, our self-awareness, our ability to distinguish right and wrong – that make a being a person. But there is a difference between human beings and animals with all of these.

Singer responds that these differences are not relevant when it comes to the important capacity that human beings and animals share, namely sentience, the basic consciousness needed to experience pleasure and pain. For a utilitarian, an act (or rule) is wrong if it produces more suffering than an alternative. Who is suffering is irrelevant. When it comes to suffering, animals should be treated as equal to people.

If the happiness of every being that can feel pleasure or pain, or can have preferences that are satisfied or not, makes a difference to whether an action is morally right or not, the problem of calculation is intensified. If it is difficult to compare the happiness of different people, it is much more difficult to compare the happiness of a person with that of, say, a pig or a bird. But this will be relevant if we ask whether we should eat meat or whether we should destroy wild habitats to make new farms to grow crops for people.

Furthermore, in response to Mill, we can argue that our inherited morality isn’t much help here. Many cultures do not take much moral notice of animals beyond prohibiting deliberately inflicting pain on them for no good reason. Apart from that, they allow treating them simply as tools to be used to make human lives happier, e.g. as food, as clothing, as beasts of burden, as objects of experiment, and so on. So we cannot use our existing moral rules, since they do not take account of the happiness of creatures other than humans, and yet we cannot calculate the effects of our actions on the happiness of non-human animals with any degree of accuracy.

Singer can respond that a new customary morality is needed. We will need to work out, together and as best we can, drawing on whatever evidence we can, the tendencies our actions have to affect the happiness of other creatures, and then create secondary principles concerning how we treat animals. It may not be perfect, but it is the best we can do.

Objection 2: Fairness and Rights

Questions of fairness, liberty and rights are questions of justice. Justice is the principle that each person receives their ‘due’. It requires that we treat equals equally, and if what someone is due depends on some quantifiable attribute (e.g. ability in some area), we should treat differences proportionally. In other words, justice is fairness. A situation is unjust, for instance, if someone has more or less than their ‘fair share’, if they are favoured or unfavoured in some way that they do not deserve. Utilitarianism requires us to take each person’s happiness into account equally. It may therefore seem that it respects fairness. But is this so?

# ‘The tyranny of the majority’

In *On Liberty*, John Stuart Mill was concerned with how, in a democracy, policies that lead to the greatest happiness for the majority could have a very negative effect on the minority. When a government is making laws, should it take into account what the majority of people want and simply overrule the interests or happiness of the minority? This may seem to be the basic principle of democracy, but if this is how laws should be made, then the majority have a form of absolute power. Suppose the majority want to enforce a system of very harsh punishment, e.g. the death penalty for theft, or outlaw homosexuality? Utilitarianism says we should do whatever brings about the greatest happiness. But should there be constraints on what the majority can do to the minority? For instance, should individuals have certain rights, e.g. to freedom of movement, freedom of thought, and freedom of expression, that protect them from such absolute power? Or is it morally permissible to remove these freedoms from some people (which would seem unfair) if it would lead to the greatest happiness overall?

Mill notes that there are two ways that the majority can exercise its power over minorities leading to a ‘tyranny of the majority’ unless its power is constrained. First, as just noted, it can do so through democratic government. For example, a democratic government could pass a law forbidding people to criticise a particular form of religious belief, or alternatively, a law forbidding them to practice it, if that is what the majority of people in society wanted. He argues that the powers of democratic government need to be limited by a respect and recognition for individual rights and liberties, but, we might think, utilitarianism has no place for these moral concepts.

Mill is also concerned about a second way the majority can exercise power, viz. through the ‘tyranny’ of social opinion, ‘the tendency of society to impose…its own ideas and practices as rules of conduct on those who dissent from them’. Everyone thinks that their way of doing things, what they like and dislike, should be the standard for everyone else. Think of the disapproval of other religious practices, of other cultures’ traditions, of homosexuality, of standards of music and taste; think how quickly people are to ‘take offence’ and think ‘something should be done about it’, and how such attitudes are communicated in the tabloid press. This disapproval, when socially expressed and endorsed, affects how people think and what they do; they are not free to think, feel and live as they please – even if there is no law preventing them from doing so.

Take, for example, the relationship between men and women. In Mill’s day, it was simply ‘understood’ that women were not equal to men. Women’s place was in the home, looking after the children. This made it very difficult for women who didn’t want to live like this – if they seriously tried to live ‘as men’, they faced strong public censure. Much of this ‘understanding’ has now, fortunately, been left behind, and a utilitarian can condemn it because women equal men in number – happiness is not maximised under sexism. But imagine now a society in which it is ‘understood’ that the place of people of some minority race in a society should serve those of the majority race. The majority of people are happy, albeit at the expense of the happiness of the minority. How could utilitarianism condemn such a practice (assuming that this policy creates the greatest happiness, i.e. any attempts to make the minority happier would lead to less happiness for the majority)?

# Rights and liberties

The obvious unfairness of the tyranny of the majority arises from the fact that act utilitarianism does not rule out any type of action as immoral. There are no constraints on pursuing the greatest happiness. For example, if torturing a child produces the greatest happiness, then it is right to torture a child. Suppose a group of child abusers only find and torture abandoned children. Only the child suffers pain (no one else knows about their activities). But they derive a great deal of happiness. So more happiness is produced by torturing the child than not, so it is morally right. This is clearly the wrong answer.

Many rights involve restrictions placed on how people can treat each other. For instance, I have a right that other people don’t kill me (the right to life). I also have a right to act as I choose as long as this respects other people’s rights (the right to liberty). One of the purposes of rights is to protect individual freedom and interests, even when violating that freedom would produce greater happiness. For example, my right to life means that no one should kill me to take my organs, even if doing so could save the lives of four other people who need, respectively, a heart, lungs, kidneys and a liver. Utilitarianism, we may object, doesn’t respect individual rights or liberty, because it doesn’t recognise any restrictions on actions that create the greatest happiness.

Some utilitarians simply accept this. We have no rights. But, they argue, as long as we consider situations realistically, then whatever brings about the greatest happiness is the right thing to do. For example, in the case of the tortured child or killing me to use my organs, in real life, other people would find out and become very upset and fearful that the same could happen to them or their children. So these actions wouldn’t lead to the greatest happiness. Counterexamples that appeal to very unlikely scenarios are unhelpful, because they have little to do with real life. In real life, act utilitarianism gives us the correct moral answer.

However, even if this is true (which we may question), the theory still implies that if it was very unlikely that anyone would find out, then it would be right to torture children (in the circumstances described). But other people finding out isn’t what makes torturing children wrong. By leaving out rights, utilitarianism misses something of great moral importance, so it can’t be the right theory of morality.

# Mill on justice

In *Utilitarianism*, Mill discusses the problem of whether utilitarianism can give a plausible account of justice, which he calls ‘the only real difficulty in the utilitarian theory of morals’.

First, he analyses what justice is and argues that at its heart, it is about the moral rights of the individual. We think of each of the following kinds of action as a violation of justice:

1. violating someone’s legal rights;
2. violating someone’s moral rights (laws are sometimes wrong, so their legal rights are not always the rights they should have in law);
3. not giving someone what they deserve, in particular failing to return good for good and evil for evil;
4. breaking a contract or promise;
5. failing to be impartial when this is required, e.g. in relation to respecting rights, what people deserve or cases of public interest;
6. treating people unequally.

What is distinctive about justice is that it relates to actions that harm a specific, identifiable individual, who has the right that we don’t harm them in this way. Duties of justice are ‘perfect’ duties. We must always fulfil them, and have no choice over when or how, because someone else has the right that we act morally. (There are other cases of wrongdoing, e.g. not giving to charity, in which no specific person can demand this of us. Instead, we have some choice in how we fulfil the obligation to help others. These are ‘imperfect’ duties.)

But why do we have the rights that we have? Mill says that ‘[w]hen we call anything a person’s right, we mean that he has a valid claim on society to protect him in the possession of it, either by the force of law, or by that of education and opinion.’ And the reason why society should protect us in this way is the general happiness. The interests that are protected as rights are ‘extraordinarily important’. They are interests concerned with security. We depend on security for protection from harm and to be able to enjoy what is good without fearing that it will be taken from us. The rules that prohibit harm and protect our freedom are more vital to our interests than any others. And so we protect these interests with rights, and these become the subject of justice. This contributes most to happiness in the long term. Hence, Mill says, ‘I account the justice which is grounded on utility to be the chief part, and incomparably the most sacred and binding part, of all morality’.

## Discussion

On Mill’s view, we only have a right if our having that right contributes to the greatest happiness in the long run. We may wonder whether the rights that we usually take ourselves to have (e.g. related to individual liberty) really do this. For example, would society be more happy if people had less freedom in some cases? This is an important debate in political philosophy.

A clearer objection is that Mill’s theory of rights doesn’t offer a strong defence of individual rights in particular cases. Suppose there is an occasion where violating my rights will create more happiness than not. As we said above, a right protects the individual’s interest against what may compete with it, e.g. the greater happiness on this occasion. Hence, my right to life prevents my being murdered to save the lives of many others. But if the ground of rights is the general happiness, this protection seems insecure. On the one hand, we have the demands of the greatest happiness, e.g. we can create more happiness if we kill one person to save five. On the other hand, we have the individual’s right, but this turns out to be just the demands of the greatest happiness as well. If my rights are justified by general utility, then doesn’t the happiness created by overriding my rights justify violating them? Utilitarianism can’t offer any other reason to respect my right in this particular instance.

Mill can respond that this approach to conflicts between rights and happiness in individual cases doesn’t understand utilitarianism in the right light. We need to consider happiness ‘in the largest sense’. Rights protect our permanent interests, and thus serve the general happiness considered over the long term. We should establish that system of rights that would bring the most happiness, and then defend these rights.

But now we can object that Mill has given up on act utilitarianism. Mill seems to recommend that we don’t look at the consequences of each act taken individually to see whether it creates the greatest happiness. He recommends that we create rights, which are a kind of rule, and enforce them even when they conflict with happiness in certain situations. Thus, he says,

[j]ustice is a name for certain classes of moral rules, which concern the essentials of human well-being more nearly, and are therefore of more absolute obligation, than any other rules for the guidance of life; and the notion which we have found to be the essence of the idea of justice, that of a right residing in an individual, implies and testifies to this more binding obligation.

When rights are involved, the right action is not the one that creates the greatest happiness, but the one that respects the right. It seems that, in the end, Mill must adopt rule utilitarianism to provide his account of rights and justice.

# Objection 3 – Partiality

Many of the things that we do to make people happy are aimed at specific other people, viz. our family and friends. We do them favours, buy them presents and generally spend our time and money on them. In other words, we are partial towards them. We are not impartial in how we treat everyone, we are not equally concerned with the happiness of everyone. But act utilitarianism argues that in our decisions, we need to consider the greatest happiness that our actions could create, and this requires us to consider the happiness of each person equally. In other words, we should be impartial.

So if act utilitarianism is right, it seems we should spend much less time with the particular people we love and more time helping people who need help, e.g. through voluntary work. Likewise, we should spend less money on the people we love and give much more money to charity. This would lead to greater happiness, because people who really need help will be made much more happy by the same amount of money or effort than people who don’t really need anything. But is this correct? Is it morally permissible to be partial or is impartiality always required of us?

There are different ways we can develop this conflict between utilitarianism and our natural partiality towards some people into an objection. For instance, we can argue that utilitarianism is too idealistic, expecting people to give priority to needy strangers over those they know and love, to be motivated by the general happiness, rather than the happiness of those they are close to. Or again, we can argue that utilitarianism misses something morally important in counting each person equally. In the abstract, each person is equal, but to me, each person does not and should not count equally. It is morally right and good (or at least, not morally wrong) to show partiality towards those people one knows and loves.

One response, which Mill gives in *Utilitarianism*, is simply to say that there are very few opportunities any of us have to benefit people ‘in general’. And so only considering and contributing to the happiness of a few people is absolutely fine, and utilitarianism does not require more. While impartiality is required in principle, in practice, we serve the greatest happiness by showing partiality. Utilitarianism is not too idealistic.

But there are two objections to this response. First, if it was true in Mill’s day that people could not often benefit people ‘in general’, that no longer seems true today. There are many charities that work around the globe and welcome volunteer fund-raisers, and the news makes us continually aware of many different causes of suffering around the world. It is perfectly possible, therefore, to dedicate much of one’s time and money to helping others ‘in general’, and there are many opportunities to do so. So it seems that utilitarianism does demand more impartiality than we usually show.

Second, Mill’s response doesn’t address the objection that utilitarianism simply fails to understand the moral importance of partiality. It is not just that partiality should be allowed. Here is an example from Michael Stocker. Suppose a woman visits a friend in hospital. The friend thanks her. She replies, ‘It was nothing, I was just doing my duty, maximising the general happiness in the world’. The friend can feel upset – the visit isn’t personal, it is just a means to create happiness. If some other action would have created more happiness, being completely impartial, the woman would have done that instead of visiting the friend. If the general happiness is the ultimate end that we should seek in our action, then we should think of our friendships as a way to maximise the general happiness. This doesn’t seem right.

Or again, consider this example from Bernard Williams. Suppose a man is in a boating accident with both his wife and a stranger. Neither can swim, and he can only rescue one. We might think that he should simply rescue his wife. But if he thinks, ‘Rescuing my wife will lead to greater happiness than rescuing the stranger’, this seems to miss the particular importance that being married has, including its moral importance. The man has ‘one thought too many’, and we (and his wife!) can object to his way of thinking about what to do.

Friendship requires that the friend is valued as the individual person that they are, and that we act out of love for them. The partiality that we show towards our friends seems to rule out a utilitarian understanding of morality. Doing something for a friend is morally good, not wrong because it fails to be impartial, nor even just ‘permitted’ as a way of maximising happiness. Furthermore, attachments of love and friendship are central to our happiness, indeed to wanting to stay alive at all. But again, these attachments motivate actions that are not impartial between everyone’s happiness. Utilitarianism fails to recognise the moral importance of partiality.

Some utilitarians have replied that morality does require us to be impartial and so it is just much more demanding than we like to think. Can you defend spending money on your friends, rather than helping others through charity, when much of the world is in poverty or at war? Rather than objecting to utilitarianism, we should side with utilitarianism in objecting to our usual, partial morality.

A different response is to say that in making the objection, we are considering utilitarianism in the wrong way. People have learned that having partial relationships is central to happiness, and so it does not maximise happiness to require people to give them up in favour of promoting the general happiness all the time.

We may object that this response gives up on act utilitarianism. Instead, it appeals to general rules about living, and considering which of these rules would maximise happiness. And this is rule utilitarianism.

# Moral integrity

Having integrity involves acting according to your own values, sticking to them especially in the face of temptation or other situations that would make it easier to do something you consider wrong. Just as our actions are guided by our concern for particular other people (partiality), they are also guided by our values. But just as utilitarianism appears to require us to set aside our partiality, it can also seem to require that we set our moral values aside in order to maximise happiness. In other words, utilitarianism attacks our moral integrity.

Bernard Williams gives the following example in *Utilitarianism*. Imagine George, who has just completed his PhD in chemistry and is looking for a job. He isn’t having much luck, and with only his wife working and small children to look after, the situation is causing a lot of stress. This is having a damaging effect on everyone, but especially the children. An older chemist says he can get George a job in a laboratory that does research developing chemical weapons. George is strongly opposed to chemical warfare and so wants to refuse the job. But his colleague points out that refusing the job will simply leave the vacancy for someone else, someone who will pursue the research with enthusiasm and so develop more chemical weapons faster and more effectively than George will.

Utilitarianism says that George should take the job. Suppose we add to Williams’ example, and say that George not only has scruples about chemical weapons, he went into chemistry in order to develop ways of counteracting chemical weapons. He came to this decision after travelling abroad, during which time he came across an awful scene in a village that had suffered a chemical attack. All this makes little difference to utilitarianism. That George would be made unhappy by taking the job is already taken into account in calculating the greatest happiness; his unhappiness in acting against his moral values is outweighed by the prevention of significant unhappiness to others.

Utilitarianism requires George to take a value he holds dear, a commitment he has made, one that bestows meaning on his life for him, and treat it as simply one preference among others because of the situation he finds himself in. It sees the situation this way: if George refuses to take the job, a consequence of his action will be that someone else will bring about significant harm (in chemical weapons research) that George can prevent by taking the job, and so George should take the job.

But can it make sense to think about George and how he chooses to live his life in this way? Why should George be responsible for what someone else does? Williams comments that

It is absurd to demand… that he should just step aside from his own project… and acknowledge the decision which utilitarian calculation requires. It is to alienate him in a real sense from his actions and the source of his action in his own convictions… It is thus… an attack on his integrity.

For each of us, our relation to what we each do is special. I am not responsible for what you do in the same way that I am responsible for what I do. But, if what you do is a consequence of what I do, utilitarianism treats them both the same. It doesn’t respect the way that my actions are expressions of who I am and the values I hold. Utilitarianism cannot understand or respect integrity.

A utilitarian such as Mill could respond that integrity is central to happiness. It doesn’t maximise happiness to require people to act against their integrity. But we may repeat that in this response, the utilitarian gives up on act utilitarianism and appeals to a rule.

# Intentions

Act utilitarianism claims that an action is right if it leads to the greatest happiness. It does not, therefore, recognise the moral value of our intentions in acting as we do. We could capture this point by saying ‘it is the thought that counts’. Whether someone intends to harm us or not – whether or not they do harm us – makes a big difference to how we respond to their action. Trying to harm someone and failing – so they are unharmed – is (usually) still blameworthy; trying not to harm someone and failing – so they are accidentally harmed – is not. But how can this be if all that matters are consequences, not intentions?

Mill discusses these points briefly in *Utilitarianism*. It is correct to say that utilitarianism considers people’s intentions as irrelevant to whether the action is morally right or not. However, that does not mean that it thinks intentions have nothing at all to do with morality. They are relevant when considering whether someone is a morally good person. And a utilitarian can say that an intention that tends to produce morally wrong actions, such as intending to harm someone, is itself a bad intention, while intentions to produce happiness are good intentions. But we need to separate the judgment of whether an action was right or wrong from the judgment of whether the intention was good or bad.

Good intentions can contribute to the greatest happiness in another way. Having good intentions one of the ‘ingredients’ of happiness. Mill argues that the desire to do good is one of those things that is desirable (good) for its own sake. For people who desire to do good because it is good, it is part of their happiness that they have this motive. Doing good is, in itself, pleasant to them. If we desire to do good, and nothing prevents us, then we also intend to do good. So good intentions are also part of a good person’s happiness. For the utilitarian, this is the best possible psychology to have. What is good is maximising happiness, and here is someone who aims at and gets happiness from maximising happiness – what could be happier!

Is Mill’s response adequate? The objection was that someone’s intentions make a moral difference to their action, e.g. that an action can be wrong because of the individual’s intentions, whatever the consequences of the action. Mill continues to deny this, and claims it only makes a moral difference to how we evaluate them as a person. Suppose someone lied to you but you saw through it – would you only think that they were a bad person or would you also think that they had done something wrong? Or again, suppose two scientists develop a genetically modified disease. One does so in order to kill people and deliberately releases it in a crowded city, wreaking havoc. The other does so in order to understand how the disease works, and takes many precautions to prevent the modified disease from escaping. But it does escape, in a crowded city, wreaking havoc. Did both scientists do equally wrong actions, or should we blame the evil scientist more than unfortunate one?

1.2. Kant’s Deontology

In this section, we discuss Kant’s claim in *Foundations of the Metaphysics of Morals* that the fundamental principle of morality is this: ‘Act only on that maxim through which you can at the same time will that it should become a universal law’. What did Kant mean by this?

To understand Kant’s moral philosophy, we need to explain a couple of terms and assumptions. First, Kant believed that, whenever we make a decision, we act on a *maxim*. Maxims are Kant’s version of intentions. They are our personal principles that guide our decisions; e.g. ‘to have as much fun as possible’, ‘to marry only someone I truly love’. All our decisions have some maxim or other behind them. Second, morality is a set of principles that are the same for everyone and that apply to everyone. Third, Kant talks of our ability to make choices and decisions as ‘the will’. He assumes that our wills are rational; that is we can make choices on the basis of reasons. We do not act only on instinct. We can act on choice, and we can consider what to choose using reasoning.

# Hypothetical and categorical imperatives

Kant calls his fundamental principle of morality the ‘Categorical Imperative’. An ‘imperative’ is just a command, a statement of what one should or ought to do. ‘Hypothetical imperatives’ are statements about what you ought to do, on the assumption of some desire or goal. They specify a means to an end. So ‘if you want to see the show, you ought to get to the theatre at least 15 minutes early’ is a hypothetical imperative. In this example, the assumed desire or goal is explicit: the imperative is presented as a conditional, with the desire described in the antecedent (‘you want to see the show’), and the command in the consequent (‘get to the theatre at least 15 minutes early’). But hypothetical imperatives can leave the assumed desire or goal implicit; e.g. ‘Eat at least five portions of fruit and vegetables a day’ (if you want to stay healthy).

Why can’t I just say ‘I want to see the show but refuse to get there early’ or ‘I want to be healthy but refuse to eat fruit and vegetables’? Why *ought* I to do these things, given what I want? Because these are the means to my end. Kant argues that willing the end *entails* willing the means. It is an analytic truth that someone who wills the end wills the means. To will an end is to will an effect. But the concept of an effect contains the concept of a cause. Hence, to will an effect, you must will the cause. The cause is the means. (It is important here that you don’t merely *want* the end, but actually will it.)

Hypothetical imperatives can be avoided by simply giving up the assumed desire or goal. Suppose I don’t want to see the show – then I don’t need to get to the theatre early. Suppose I don’t want to be healthy – then the imperative to get my ‘five-a-day’ doesn’t apply to me. (Of course, it is odd not to want to be healthy, and we may wonder if I really do not want to be healthy. Perhaps I do want to be healthy, but I can’t be bothered. If this is the case, I want to be healthy, but I don’t will it.) In other words, it is possible to ‘opt out’ of a hypothetical imperative.

This isn’t true of morality, we usually think. Moral duties are not hypothetical. They are what we ought to do, full stop. They are your duty regardless of what you want. They are ‘categorical’. Kant has also argued that moral duties aren’t a means to some further end, because what makes an action good is that it is willed by the good will. All categorical imperatives – our moral duties – are derived from one, *the* Categorical Imperative: ‘Act only on that maxim through which you can at the same time will that it should become a universal law’.

How are categorical imperatives possible? Why is there something that we ought to do, regardless of what we want? Kant argues that moral duties depend just on our being rational. We need to understand further just what this means.

# Contradiction in conception and contradiction in will

There are two different ways in which we could fail to be able to will our maxim to become a universal law.

‘Contradiction in conception’: the situation in which everyone acted on that maxim is somehow self-contradictory. Suppose you want a gift to take to a party, but you can’t afford it, so you steal it from the shop. Your maxim is something like: ‘To steal something I want if I can’t afford it’. This can only be the right thing to do if everyone could do it. However, if we could all just help ourselves to whatever we wanted, the idea of ‘owning’ things would disappear. Now, by definition, you can’t steal something unless it belongs to someone else. Stealing presupposes that people own things. But people can only own things if they don’t all go around helping themselves whenever they want. So it is logically impossible for everyone to steal things. In other words, it is inconceivable – a contradiction in conception – for everyone to steal things. We can’t conceive of the maxim ‘To steal something I want if I can’t afford it’ being a universal law, so we can’t rationally will it to be a universal law. And so stealing (at least stealing just because one wants something) is wrong.

‘Contradiction in will’: this is more difficult to understand. The maxim is not self-contradictory when universalised, but there is another way in which we cannot rationally will it. Consider a refusal to help other people, ever. It *is* logically possible to universalise the maxim ‘not to help others in need’. The world would not be a pleasant place, but this is beside the point. Kant does *not* claim that an action is wrong because we *wouldn’t like* the consequences if everyone did it (many philosophers and students have misinterpreted Kant on this point). His test is whether we can rationally will that our maxim be a universal law. Willing and wanting (or liking) are different. Someone can want something that they don’t will – they don’t choose to act on their desire, e.g. such as cheating on their husband or wife with someone they find very attractive. And someone can will something they don’t want, such as going to the hospital for surgery. Kant is concerned with willing not wanting. He argues that we *cannot will* that no one ever help anyone else. How so?

P1. A will, by definition, wills its ends (goals).

P2. As we said above, to truly will the ends, one must will the necessary means.

C1. Therefore, we cannot rationally will a situation in which it would be impossible for us to achieve our ends. To do so is to cease to will the necessary means to one’s ends, which is effectively to cease to will any ends at all. This contradicts the very act of willing.

P3. It is possible that the only available means to our ends, in some situations, involves the help of others.

C2. We cannot therefore will that this possibility is denied to us.

C3. Therefore, we cannot will a situation in which no one ever helps anyone else.

# Morality and reason

As the contradiction in conception and contradiction in will show, disobeying the Categorical Imperative involves a self-contradiction, according to Kant. He argued that it is not just morally wrong to disobey the Categorical Imperative, it is also irrational. Through the Categorical Imperative, *reason* both determines what our duties are and gives us the means to discover them. Furthermore, we intuitively think that morality applies to *all and only* rational beings, not just human beings.

In Douglas Adams’ *The Hitchhiker’s Guide to the Galaxy*, Arthur Dent protests to the Vogons, rational aliens who are going to destroy the Earth, that what they are doing is immoral. Dent’s protest makes sense, even though he isn’t protesting about the actions of human beings. But morality doesn’t apply to beings that can’t make rational choices, such as dogs and cats (pets misbehave; they don’t act *morally wrongly*).

With this link, we can explain the nature of morality in terms of the nature of reason. Morality is universal, the same for everyone; so is reason, says Kant. Morality and rationality are categorical; the demands to be rational and moral don’t stop applying to you even if you don’t care about them. Neither morality nor rationality depend on what we want.

# The second formulation of the Categorical Imperative

Kant gives a second formulation of the Categorical Imperative, known as the Formula of Humanity: ‘Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end’. Why does he say this, and what does it mean?

Let us return to the idea of the good will. Only the good will is good without qualification. Another way of saying this is that it is the only thing of unconditional value. Everything else that is valuable depends, in some way, on the good will. For instance, intelligence is valuable for all sorts of purposes. In other words, it is valuable as a means to an end. Its value, then, depends on the value of its end. What gives its end value? We do, says Kant. Something is only an end if it is adopted by a will. It is our adopting something as an end that gives it value. Because I have desires and purposes, various things in the world are valuable *to me*.

So far, value is subjective. What is valuable is valuable because it is valuable to someone; and what is valuable to me may not be valuable to you. However, this does not apply to other people (or rational beings generally). Your value is not simply your value *to me* as a means in relation to some purpose or desire I have. It is not even your value to you (you might have very low self-esteem, and wrongly underestimate your value). We have ‘intrinsic worth’, which Kant identifies as ‘dignity’. What gives us this dignity is our rational will. The will has unconditional value *as the thing which gives value to everything else*. So in the second formulation above, by ‘humanity’, Kant means our ability to rationally determine which ends to adopt and pursue.

Kant says that because people are ends in themselves, we must always treat them as such, and never ‘simply’ as a means. Note that he does not say we cannot use people as a means, but that we can’t use them *only* as a means. We rely on other people in many ways as a means of achieving our own ends; e.g. people serving me in a shop are a means of getting what I want to buy. What is important, says Kant, is that I also respect them as an end.

To treat someone simply as a means, and not also as an end, is to treat the person in a way that undermines their power of making a rational choice themselves. It means, first, that we should appeal to other people’s reason in discussing with them what to do, rather than manipulating them in ways they are unaware of. Coercing someone, lying to them or stealing from them all involve not allowing them to make an informed choice. If they are involved in our action in any way, they need to be able to agree (or refuse) to adopt our end as their own.

Second, treating someone as an end also means leaving them free to pursue the ends that they adopt. The value of what people choose to do lies in their ability to choose it, not just in what they have chosen. So we should refrain from harming or hindering them. This is to respect their rationality. Third, someone’s being an end in themselves means that they are an end for others. We should adopt their ends as our own. What this means is that we should help them pursue their ends, just as we pursue our own ends. In other words, the second formulation requires that we help other people. This should be one of our ends in life.

# Universalisability and morality

Is Kant right to think that acting on maxims that are universalisable is morally right or permissible, while acting on maxims that are not universalisable is morally wrong? Are there counterexamples? Let’s start by asking whether there be a case of acting on a universalisable maxim that is morally wrong. We might think that this is just a matter of phrasing the maxim cleverly. In the example of stealing the gift above, I could claim that my maxim is ‘To steal gifts from large shops and when there are six letters in my name (Daniel)’. Universalising this maxim, only people with six letters in their name would steal only gifts and only from large shops. The case would apply so rarely that there would be no general breakdown in the concept of private property. So it would be perfectly possible for this law to apply to everyone.

Kant’s response is that his theory is concerned with my actual maxim, not some made-up one. It is not actually part of my choice that my name has six letters, or perhaps even that it is a gift I steal. If I am honest with myself, I have to admit that it is a question of my taking what I want when I can’t afford it. For Kant’s test to work, we must be honest with ourselves about what our maxims are.

Can we find a counterexample of this kind? Suppose I am in dire straits. I really need money to get food and shelter, and the situation is growing urgent. But I am too proud to ask people for help. So I con them instead: I borrow money on the promise of repaying it, but I don’t intend to keep my promise. I wouldn’t do this unless things were desperate. Is my maxim universalisable? It seems so. If, as a matter of law, everyone made promises they didn’t intend to keep whenever they wanted something, that would be impossible. People would no longer believe promises; and you can’t make a promise unless someone accepts it! But my maxim is much more specific, because it is ‘to make a promise I don’t intend to keep rather than ask for help, but only in the face of such desperate circumstances’. This can be universalised, it seems, as it wouldn’t occur often enough for promise-making to become impossible.

But now, is what I do wrong? If we think it is not, then this example is no counterexample – my maxim can be universalised, and my act is not wrong. But if conning people in this situation, rather than asking for help, is wrong, then this is an action that is wrong, and yet the maxim is universalisable. Kant’s Categorical Imperative fails to give us the right answer. It is not always right to do something where the maxim is universalisable.

Another possible counterexample is furnished by one of Kant’s own examples in ‘On a Supposed Right to Lie from Altruistic Motives’. Suppose someone comes to your house to seek refuge from someone who wants to murder them. Soon after they have hidden, the would-be murderer arrives and asks you where they are. Even in this situation, Kant says, you should not lie. Lying is always wrong, because we cannot universalise the maxim to deceive people. Most people would disagree, and argue that lying in such a situation is the right action. But perhaps Kant is wrong about our maxim in this case. Perhaps the maxim is to tell a lie to save a life can be universalised.

Are there any clearer counterexamples of maxims where acting on them is not morally wrong, but the maxim is not universalisable? Say I am a hard-working shop assistant, who hates the work. One happy Saturday I win the lottery, and I vow ‘never to sell anything to anyone again, but only ever to buy’. This is perhaps eccentric, but it doesn’t seem morally wrong. But it cannot be universalised. If no one ever sold things, how could anyone buy them? It is logically impossible, which makes it wrong according to Kant’s test. So Kant’s Categorical Imperative again gives us the wrong answer. It is not always wrong to do things which require other people do something different.

# Morality is a system of hypothetical imperatives

Is Kant right to think that failing to act in accordance with the Categorical Imperative is a failure of reason as well as morally wrong? Are there, in fact, any categorical imperatives, rules we must follow on pain of being irrational if we don’t? In ‘Morality as a system of hypothetical imperatives’, Philippa Foot argues that there are not, and that what it is rational for someone to do depends on what they want.

An imperative for Kant is something that ought to be done. Hypothetical imperatives state that you should or need to do the action as a means to something you want. Here ‘want’ has a wide meaning, covering not just passing or occasional desires, but also your long-term projects and plans. For example, if you are committed to getting a good education, then you should study hard, even when you don’t feel like it. Categorical imperatives present the action as some something you should or must do, full stop, ‘without regard to any other end’. Kant argues that they are ‘objective’, that acting on them is a matter of being rational rather than fulfilling a subjective desire.

Foot notes that we do commonly contrast moral judgments with hypothetical imperatives in this sense. We find two uses of ‘should’ or ‘ought’ in how we use language. In the first use, if we discover the person doesn’t want what the imperative assumes, or we discover the action isn’t a suitable means, then we no longer say they should act on it. For example, ‘you should take the third left if you are going to the restaurant’: if you are not going to the restaurant, or the third left is a dead-end, then we no longer say that you should take the third left. But we also use the words ‘should’ and ‘ought’ when we don’t withdraw our claim that ‘you should do x’. ‘You shouldn’t lie’, ‘But I don’t care about the truth, I really want to trick him’, ‘That’s irrelevant; you shouldn’t lie’. You can’t rebut or escape the requirement to act just by showing that it doesn’t help you get what you want.

However, so far, this point is only about language, and it isn’t enough to show that Kant is right that moral judgments are categorical in the sense he means. To see this, think of the rules of etiquette or the rules of a club. For example, in the UK, handshakes should be brief (so Debrett’s guide to etiquette tells us). If… what? The imperative doesn’t mention something you want. We might try to spell it out, e.g. adding ‘if you want to fit in’ or ‘if you want people to think well of you’, but this isn’t how etiquette works. Even if someone doesn’t care what others think, it is still a breach of etiquette for them not to release another person’s hand after shaking it. Likewise, in Foot’s (now old-fashioned) example, if the club rules say, ‘Do not take ladies into the smoking room’, there isn’t a hidden assumption ‘if you want to remain part of the club’. Suppose someone doesn’t want to remain part of the club, thinking it fusty and sexist, and he will quit tomorrow for good. Is he now allowed to take ladies into the smoking room? No. In these examples, we don’t withdraw the ‘should’ depending on what someone wants.

These are examples of non-hypothetical imperatives. Clearly, they are not unconditional or inescapable in the sense that Kant thinks moral judgments are. They are not categorical in Kant’s sense because by themselves they don’t give us a reason to act. Whether you have reason to observe the rules of etiquette or the rules of the club will depend on what you want. If you don’t like the rules of the club, don’t join – nothing wrong in that.

Moral judgments are also non-hypothetical imperatives in this sense. But this fact does not yet show that they are categorical in the sense of giving everyone and anyone a reason to act in accordance with them. To show this would be to show that immorality is irrational, which is just what Kant argues. But, claims Foot, this is because Kant assumes that acting immorally involves disregarding a rule that you have accepted (e.g. that no one should lie) or again that it is inconsistent to want other people to act in a way you don’t intend to (e.g. that they should tell the truth while you lie). But this is simply not so. ‘Irrational actions are those in which a man in some way defeats his own purposes’, and acting immorally need not involve this (although it may, e.g. by making enemies of people you may later need as friends).

Why do we think that the rules of morality are categorical when the rules of etiquette are not, even though both are non-hypothetical? Foot argues that the answer lies in our feelings about morality. The ‘binding force’ of morality is simply the feeling that moral judgments are inescapable. And our feelings are the result of how moral rules are taught. The rules of morality are taught and enforced much more stringently than the rules of etiquette. As a result of how we as children are trained to behave in morally right ways, we feel that we ‘must do’ what is morally right, whatever our desires or plans. There is no other meaning we can sensibly give to the idea that morality is ‘categorical’.

Foot recognises two possible objections to her view. First, if she is right, then what does ‘acting out of duty’ amount to? If moral judgments are not categorical, it seems that doing what is right ‘because it is right’ no longer gives us a reason to act. Foot’s response is that Kant is mistaken in thinking that the motive of duty was the only morally good motive.

We genuinely care about others’ good quite apart from thoughts of duty. We can understand ‘doing the right thing because it is the right thing to do’ as being motivated by morally good concerns. This looks like moral action on the basis of hypothetical imperative, e.g. you are helping because you want to help. This isn’t wrong, but ‘wanting to help’ needn’t be a passing desire. A virtuous person is one who is dedicated to moral ends, not someone who acts morally just so long as they ‘feel inclined’.

Second, doesn’t Foot’s view undermine morality? In particular, what can we say to people who simply don’t care about morality (‘amoralists’)? Isn’t it true they ought to care? And isn’t this a contrast with the case of someone who doesn’t want to join the club? Foot responds that amoralists could accept that the moral ‘ought’ is non-hypothetical, but still not think it gives them a reason to live by moral rules. Amoralists take themselves to have no reason to be moral. We can say that they may well be mistaken, and could spoil their own lives. But there is no more that we can say than this, for the moral ‘ought’ has no magical force to give everyone a reason to be moral, irrespective of what they want in life.

# Conflicts between duties

Kant argues that our moral duties are absolute. A duty is absolute if it permits no exceptions. Nothing can override a moral duty, because it is categorical. All other ends have their worth in relation to the good will. But the good will is motivated by duty.

This causes problems in cases in which it seems that two absolute duties conflict with each other. Should I break a promise or tell a lie? Should I betray a friend to save a life? If I am faced with a situation in which I must do one or the other, then Kant’s theory implies that whatever I do must be wrong.

One response is to say that a real conflict of duties can never occur. If there appears to be a conflict, we have misunderstood what at least one duty requires of us. If duties are absolute, we must formulate our duties very, very carefully to avoid them conflicting. Kant himself thought that some of our duties are very straightforward; e.g. our duty not to lie is simply that – never lie. But you can believe the rest of Kant’s theory and not accept his view that duties are simple like this. For example, you could argue that ‘don’t lie’ isn’t a duty. Our duty could be something like ‘don’t lie unless you have to lie to save a life’. There will always be some maxim you can act on which you will be able to universalise. So it will always be possible to do your duty.

We can object that it is more realistic simply to say that (most) duties are not absolute. For instance, there is a duty not to lie, but it may be permissible to lie in order to save someone’s life. Less important duties can ‘give way’ to more important ones. In cases of conflict, one will give way and no longer be a duty in that situation. This understanding is at odds with Kant’s theory of morality. His whole analysis of duty is that it is categorical. It is difficult to see how his understanding of why morality is rationally required could allow that duties can give way to each other.

# The view that consequences of actions determine their moral value

Utilitarians object that Kantian deontology is confused about moral value. If it is my duty not to murder, for instance, this must be because there is something bad about murder. But then if murder is bad, surely we should try to ensure that there are as few murders as possible. If I know that unless I kill someone deliberately, many people will die, how can I justify not killing that person? Surely it is only my duty not to kill because death is bad. So I should prevent more deaths, and so in this case, I have a duty to kill, because I would be killing in order to save lives. What makes a will good is that it wills good ends.

Kant’s response is that there are no ends that are good without qualification, even happiness. Apart from the ‘good will’ itself – the will that only acts on universalisable maxims – nothing is unconditionally good. For instance, intelligence and self-control are good – but they can enable someone to do clever or difficult bad things, if that is what they choose. Power can be good, but it depends on what use we put it to. Nor is happiness good without qualification. If someone is made happy by hurting others, their happiness is morally bad. So we evaluate happiness by morality. Having a morally good will is a precondition to *deserving* happiness. So utilitarianism does not provide the right analysis of the good will.

But the disagreement goes deeper. Utilitarianism understands all practical reasoning – reasoning about what to do – as means–end reasoning: it is rational to do whatever brings about a good end. The utilitarian thinks it is just obvious that if something is good, more of it is better, and we ought to do what is better. Kant disagrees and offers an alternative theory of practical reasoning. Means–end reasoning is appropriate for hypothetical imperatives, but this is not all there is to practical reason. It is also irrational to act in a way that not everyone could act in. If rationality were only about means–end reasoning, then we couldn’t say that any ends – such as other people – are obligatory. Morality becomes hypothetical. You only ought to do your duty if you want to be morally good. This treats morality like just another desire or purpose which we may or may not have.

Utilitarians respond that happiness is the only desirable end. But, once again, Kant has argued that happiness is not always good.

# The value of certain motives

Kantian deontology does not require us to be impartial between our friends/family and people we don’t know. While we are required to help others, we are not required to be completely impartial or maximise happiness. There is no contradiction in maxims that show partiality to some people. (And there is no contradiction in a maxim which aims to help others but not maximise happiness.) However, can Kant’s theory recognise and explain the moral worth of motives involved in relationships of love and friendship?

Kant makes the motive of duty, acting out of duty, doing your duty because it is your duty, the only motive that has moral worth, and says that doing something good for someone else because you want to is morally right, but not morally good. But consider this example from Michael Stocker. Suppose a woman visits a friend in hospital. The friend thanks her. She replies, ‘It was nothing, I was just doing my duty’. If her motive really is simply to do what is morally right, then her friend can object. Kant seems to say that we have to want to benefit people because it is our duty to so, not because we like them. But surely, if I do something nice for you because I like you, that is a morally good action.

This applies as much outside relationships of partiality. I may act to help a stranger, moved by kindness, sympathy and compassion. My action may well be in accordance with duty, but because I am motivated by my feelings and not by a concern to do the right thing because it is the right thing to do, Kant would say that my action has no moral worth. My feelings are instrumentally valuable, because they motivate me to act in accordance with duty. But they are not themselves morally valuable. And yet much of the time, we do good things because we feel warmly towards the people we benefit. We can object that putting the motive of duty above feelings as the source of good action is somehow inhuman.

Kant can respond that he is not trying to stop us from being motivated by our feelings. His point is that, when we are choosing what to do, how we feel should not be as important as what it is morally right to do. Our feelings shouldn’t decide the matter, our motive to do what is morally right should. But when you do something for a friend, should you think ‘I’ll do this because he is my friend; and it is morally right to do so’? Consider this example from Bernard Williams. Suppose a man is in a boating accident with both his wife and a stranger. Neither can swim, and he can only rescue one. We might think that he should simply rescue his wife. But if he thinks, ‘She’s my wife and it is morally permissible that I rescue her’, this seems to miss the particular importance that being married has, including its moral importance. The man has ‘one thought too many’, and we (and his wife!) can object to his way of thinking about what to do. His commitment to his wife means that he should stop at ‘She’s my wife’.

Perhaps Kant can reply that you don’t actually need to have such a thought. His theory, after all, is how we can tell whether something is right or wrong, not how we should actually think all the time. So we can say that to be morally good, you only need to be willing to refuse to help your friend if that involved doing something morally wrong. And likewise for the man and his wife.

1.3. Aristotelian Virtue Ethics

Some normative ethical theories, such as utilitarianism and Kantian deontology, focus on morally *right actions* – what is the right thing to do, and why? On these views, to be a good person is to be motivated to do morally right actions. By contrast, virtue ethics starts with what it is to be a *good person*. From this, it then derives an account of what a morally right action is, which it understands in terms of what a good person would do. An important claim of virtue ethics is that there is more to the moral life than actions.

Some form of virtue ethics or other has been the ‘default form’ of ethical theory in Western philosophy until the last few hundred years. Its ‘classical’ version is stated most clearly by Aristotle, and other forms developed by rejecting or adding to some elements of Aristotle’s theory. In this section, we discuss Aristotle’s concept of *eudaimonia*.

# The good for human beings

Aristotle begins the *Nicomachean Ethics* with the question ‘What is the good for human beings?’ What is it that we are aiming at, that would provide a successful, fulfilling, good life? Our different activities aim at various ‘goods’. For example, medicine aims at health; military strategy aims at victory. For any action or activity, there is a purpose (a ‘why’) for which we undertake it – its end. An analysis of the purposes for which we do things is an analysis of what we see to be ‘good’ about them. An answer to ‘Why do that?’ is an answer to ‘What’s the point?’ – and ‘the point’ is what is worthwhile about doing that.

Now, complex activities, such as medicine, have many component activities, e.g. making pharmaceuticals, making surgical implements, diagnosis, etc. Where an activity has different components like this, the overall end (health) is better – ‘more preferable’ – than the end of each subordinate activity (successful drugs, useful implements, accurate diagnoses). This is because these activities are undertaken for the sake of the overall end.

We undertake actions and activities either for the sake of something further or ‘for their own sake’. Suppose there is some end for whose sake we do everything else. Suppose that this end we desire for its own sake, not the sake of anything else. Then this end would be the good for us. As Julia Annas notes in her article ‘Virtue ethics’, in thinking about why we do what we do, we end up thinking about how to live one’s life well as a whole.

# Eudaimonia

People generally agree, says Aristotle, that this is ‘eudaimonia’. What does he mean by this?

Eudaimonia is the good for a human life. It is often translated as ‘happiness’ but Aristotle says it is ‘living well and faring well’. We have some idea of what it is when an animal or plant is living and faring well – we talk of them ‘flourishing’. A plant or animal flourishes when its needs are met in abundance and it is a good specimen of its species. Gardeners try to enable their plants to flourish; zookeepers try to enable the zoo animals to flourish. So eudaimonia is ‘the good’ or the ‘good life’ for human beings as the particular sort of being we are. To achieve it is to live as best a human being can live.

There are a number of contrasts we can draw with our usual idea of ‘happiness’.

1. We can talk of people being happy as a psychological state, and in particular – perhaps a result of the influence of utilitarianism – we think of it as pleasure. But eudaimonia is not a state of mind. It characterises an activity – the activity of living. A good life is one that realises the full potential that a human life has.
2. Eudaimonia is not something subjective, but objective. To say someone is or was eudaimon is to make an objective judgement about their life as a good human life. It is not to say anything (directly) about their state of mind; nor is it a judgement the person themselves has any special authority over. By contrast, if someone says they are happy or unhappy, it is difficult to correct them or know better.
3. Eudaimonia is not something easily changed. It does not come and go as happiness (in the usual sense) can. For it is an evaluation of a person’s life (a life lived well) as a whole. This is a very stable judgement.

However, we still don’t know just what eudaimonia is – what sort of life is a good or flourishing life? Aristotle notes that people disagree on whether it involves pleasure, wealth, honour, or something else again. But, says Annas, if we start from the idea that it characterises the activity of living one’s life, it can’t be about passive states of mind, such as pleasure. Another reason it can’t be just pleasure per se, Aristotle argues, is because we share pleasure with animals and we’re after the good for human beings.

It also can’t be about money or wealth. First, notes Annas, having wealth isn’t an activity. Second, if eudaimonia is a final end, then it can’t be an instrumental good. But money is only useful as a means to an end, it isn’t an end in itself. Aristotle argues that it can’t be honour either, since to have honour, others must honour you. What is it you want to be honoured (recognised, rewarded, praised) for? Whatever the answer, achieving that must be what is good.

Aristotle briefly raises the suggestion that the wise person wants to be honoured for their virtues. (We’ll consider what a virtue is below.) But just having virtues, e.g. courage or intelligence, can’t be enough for a good life, for two reasons. First, you can have virtue while asleep. Such inactivity isn’t our end in life. Second, having virtue is compatible with suffering great misfortune in life. But this isn’t a good life either. So we still don’t know yet what eudaimonia is.

# Final ends

Is there such a thing as the good for human beings? Given that we think pleasure, honour, or again, knowledge, are all good, how could eudaimonia be the good, our only good?

Call an end that we desire for its own sake a ‘final’ end. We can’t give some further purpose for why we seek it. If there is just one end for the sake of which we do everything else, that is the good. If there is more than one end, there are various final ends, each of which is good. If pleasure, honour and knowledge are final ends, doesn’t that show that eudaimonia is not our only good?

Not yet. Some final ends we might seek both for their own sake and for the sake of something else. Everything that we pursue for its own sake – such as pleasure, knowledge, honour, and so on – we also pursue for the sake of eudaimonia, of living a good life.

How can we pursue something both for its own sake and for the sake of eudaimonia? The solution is to distinguish between ‘external means’ and ‘constitutive means’. We usually think of the relation between means and end as an instrumental relation; i.e. that performing the means achieves the further, independent end. Think about having a good holiday. Suppose you have to get up very early in order to catch the plane. You do this in order to have a good holiday, but it isn’t part of having a good holiday. Getting up early is an external means to the end. But there is also another relation between means and ends, a constitutive relation. Later on, you are lying on the beach in the sun, listening to your favourite music. Are you doing this ‘in order’ to have a good holiday? Not in the same sense. This just is having a good holiday at the moment. Lying on the beach is a constitutive means to the end of having a good holiday. Having a good holiday is not something ‘further’ or additional that you achieve by lying on the beach. In these circumstances, here and now, it is what ‘having a good holiday’ amounts to.

Final ends are constitutive parts of eudaimonia. For example, we can pursue knowledge for its own sake and pursue it for the sake of living well if we believe that acquiring knowledge is part of the good life.

Everything we do, says Aristotle, is done for the sake of living and faring well. By contrast, we never want to live and fare well in order to achieve some other end. If there is a final end which we never seek for the sake of anything else, but only ever for its own sake, this will be a final end ‘without qualification’. Annas comments that if eudaimonia is our final end, then it must be ‘complete’ in just this sense.

A further reason for thinking eudaimonia is our only good is that the good should be self-sufficient; i.e. it makes life desirable on its own. Eudaimonia is the most desirable thing, and we can’t make it more desirable by adding something else to it. In fact, given what we’ve just said, to add some other goal, e.g. knowledge, to eudaimonia is just to make that other thing part of your eudaimonia. Eudaimonia is the only self-sufficient good.

# ‘Function’ and ‘virtue’

Having established the relation between eudaimonia and other goods, we need to think again about what eudaimonia is. So Aristotle embarks on an analysis of eudaimonia in terms of the idea of ergon. This is often translated ‘function’, but as with translating eudaimonia as happiness, this is misleading. The ergon of a thing can be its function – the ergon of an eye is to see – but a more general account would be the ‘characteristic form of activity’ of something. ‘Function’ here is better understood in relation to ‘functioning’ rather than ‘purpose’.

The ‘characteristic activity’ of something provides an insight into what type of thing it is (otherwise in what sense would the activity be ‘characteristic’?). It thereby provides an evaluative standard for that thing: something is a good x when it performs its characteristic activity well. If the ergon of a knife is to cut, a good knife cuts well; a good eye sees well; a good plant flourishes (it grows well, produces flowers well, etc., according to its species).

In order to fulfil its ergon, a thing will need certain qualities. An arête is a quality that aids the fulfilment of a thing’s ergon. It can be translated generally as an ‘excellence’, or more specifically, a ‘virtue’. So sharpness is a virtue in a knife designed to cut. Good focus is a virtue in an eye.

## The function argument

Aristotle applies this entire account to human beings. Virtues for human beings will be those traits that enable them to fulfil their ergon. So, first, what is the ‘characteristic activity’ of human beings? At the most general level, we are alive. But this isn’t distinctive of just us. So we shouldn’t identify ‘life’ as our characteristic activity. We are a type of animal, rather than plant. We are conscious, have sense perception, etc. But again, we share this with many animals. But we want to know what the good for human beings, distinctively, is.

A human life is distinctively the life of a being that can be guided by reason. We are, distinctively, rational animals. Many commentators misunderstand Aristotle to be claiming that reasoning is our ergon. But Aristotle makes a deeper point – what is characteristic of us is that whatever we do, we do for reasons. All our activities – not just ‘reasoning’ – are, or can be, guided by reasons. Being guided by reasons is, of course, a matter of our psychology, and so Aristotle talks of the activity of the soul (psyche).

Now, we said above, that a good x (eye, knife, etc.) is one that performs its characteristic activity well, and that it will need certain qualities – virtues – to enable it to do this. Our ergon is living as a rational animal, i.e. living in accordance with reason, and the virtues of a human being will be what enables us to do this. To fulfil our ergon and live well, we must be guided by the ‘right’ reasons – good reasons, not ‘bad’ reasons. So eudaimonia consists in the activity of the soul which exhibits the virtues by being in accordance with (‘good’ or ‘right’) reason. Eudaimonia is living a life in which one exercises the virtues. Finally, we must add – as noted earlier – that this must apply to a person’s life as a whole. A day or even a year of living well doesn’t amount to a good life.

# Testing the analysis

It is worth double-checking that this is a plausible account of eudaimonia. Aristotle argues that it is indeed consistent with other things we want to say about what is good for human beings.

1. There are three types of thing that are good for us – goods of the mind (e.g. intelligence, courage, etc.), goods of the body (e.g. strength, health, etc.) and ‘external’ goods (e.g. wealth, food, etc.). People generally agree that the goods of the mind are worth more than the others. We often think of the others as additional to, but not comprising, a good life. This agrees with the analysis; eudaimonia centrally concerns goods ‘of the soul’.
2. We have said that eudaimonia is living well. The analysis agrees, and spells out what it is to live well.
3. We can return to the suggestions that eudaimonia involves virtue, pleasure and wealth, and now explain the truth in each.
   1. Virtue: as we said, to possess virtue is not enough; eudaimonia requires that one acts on it as well. The employment of virtues and the achievement of good purposes are better than simply having the virtues.
   2. Pleasure: people find pleasant whatever it is that they love. A virtuous person loves living virtuously – you shouldn’t call someone ‘just’, for instance, if they dislike doing what is just. But that means that the life of the virtuous person will also be pleasant. Eudaimonia is therefore both good and pleasant.
   3. Wealth: in order to live virtuously (e.g. to be generous), we will also need a certain amount of external goods. And so, enough good fortune is needed for a fully good life.

Aristotle then raises a puzzle. If eudaimonia relates to the whole of someone’s life, then can you call someone eudaimon while they are still alive? Their life is not yet finished – something terrible may yet happen that would lead us to say that theirs was not a good life. On the other hand, it is absurd to say that they are eudaimon after they have died. We could say, once they are dead, that they were eudaimon, but then it is strange that we cannot say that they are eudaimon before they have died.

Aristotle’s solution is to say that fortunes change, but living virtuously has a much greater permanence. A virtuous person deals with bad fortune in the best possible way, so only very rarely and through terrible circumstances, can someone virtuous fail to lead a good life. Now we understand that virtue is central to leading a good life, we can call someone who is virtuous ‘eudaimon’ while they live, if they have sufficient external goods.

# Eudaimonia and pleasure

In his account of eudaimonia, Aristotle emphasises the importance of virtue and reason. But many people think that pleasure is central. What place does Aristotle give to pleasure in his account?

## Is pleasure good?

Aristotle claims that pleasure is good, and that eudaimonia involves pleasure. To defend his view, he needs to answer objections that reject the goodness of pleasure, and to clarify just how and when pleasure is good.

1. Objection: The temperate person avoids pleasure.

Reply: Not true. What the temperate person avoids is an excess of certain bodily pleasures.

1. Objection: The practically wise person doesn’t seek pleasure, but only avoids pain.

Reply: Not true. The practically wise person does seek pleasure, but in accordance with reason. Furthermore, the fact that they avoid pain (in accordance with reason) shows that pleasure is good. As pain is bad and to be avoided, the contrary of pain, pleasure, is good and to be pursued.

1. Objection: Pleasure interferes with thought.

Reply: Not true. The pleasures of thinking don’t interfere with thinking, but assist it. It is pleasures that arise from other sources that interfere with thinking. It is generally true of pleasurable activities that each interferes with the others.

1. Objection: Not all pleasures are good, for example bodily pleasures or taking pleasure in something bad or disgraceful. (e.g. voyeurism)

Reply: If we say bodily pleasures are not good, then how can we explain that their opposite, bodily pains, are bad? It is only excess of pleasure here that is bad. Disgraceful pleasures are not good, agreed. To explain this, we could say any of three things:

* 1. Disgraceful pleasures are not really pleasures, but only pleasant to bad people. All real pleasures are good, though.
  2. The kind of pleasure involved in something disgraceful is a pleasure (e.g. looking at an attractive naked body), and so it is good in general. But such pleasure is not good when it is caused by or involves something disgraceful (such as an intrusion on privacy).
  3. Pleasures are of different kinds, and only some pleasures are good. We will look further at this below.

Do we have any positive reasons for thinking that pleasure is good? Aristotle considers four arguments from another philosopher, Eudoxus, for the claim that pleasure is the only good. He argues that Eudoxus is right that pleasure is a good, but not that it is the only good. (We will see further arguments from Aristotle in the next section.)

1. Every creature aims at pleasure. This is a good indication that it is, for each thing, the good. And what is good for all things is the good.

Aristotle agrees that this is the strongest reason for thinking that pleasure is good. However, he argues that pleasure is not the only thing that we aim at, it is not our only end. There are other things which we seek out, such as seeing, knowing, being virtuous, that we would seek out even if they brought us no pleasure. The pleasure they bring is not why we seek them. They are not simply a means to pleasure.

With the next three arguments, Aristotle agrees that they show that pleasure is good, but not that it is the only good.

1. Everything avoids pain, so its contrary, pleasure, is good.
2. We choose pleasure for its own sake, not for some further purpose.
3. Adding pleasure on to any good makes it more desirable. So, we should conclude that pleasure is good, but not the only good.

## Pleasure, virtue and function

What is pleasure? We naturally think of it as a kind of subjective feeling, which we can only define by how it feels. But Aristotle argues that it is the unimpeded activity of our faculties.

This is a very difficult claim to understand, but we can start by thinking about being ‘in the zone’, as we say now. Start with the activities of the senses, such as seeing. Pleasure in the activity of a sense is caused most when that sense is at its best (e.g. when you can see well) and active in relation to its ‘finest’ object.

Aristotle doesn’t define this, but we can think of it as something on which we can really exercise that sense. So with vision, this is something that is (at least) interesting to look at, that we can explore and engage with through sight. Works of art and beautiful landscapes might provide examples. Looking at such things gives us (visual) pleasure. The same can be said of activities of thought – there is pleasure here in grappling with something that exercises our thought, but which doesn’t impede it, e.g. through being too difficult to understand. We can extend this analysis to all our activities.

But pleasure is not something simply caused by, and separate from, such unimpeded activity. It ‘completes’ the activity. It is part of it, not a separate end, nor a state produced by the activity, as deliberating might produce a decision or looking might produce finding. The pleasure is in the activity itself and intensifies and supports it. Thus, when we enjoy an activity, we throw ourselves into it, and we enjoy it less if our attention is distracted.

If this is the correct analysis of what pleasure is, we can explain how pleasures can be good or bad, and how they relate to virtue and eudaimonia. Each kind of activity – eating, thinking, running, listening to music – has a corresponding kind of pleasure. So there are different kinds of pleasure. A pleasure is good when the activity that produces it is good and bad when the activity is bad.

Aristotle claims that different animals have different characteristic activities, and so they enjoy different pleasures. The pleasures that are most suited to human beings are, therefore, those that relate to our characteristic activity, namely living in accordance with reason. Now, it is the virtuous person who has the traits and the practical wisdom that enable them to perform this characteristic activity and this constitutes the good life for human beings. So what is ‘truly’ pleasant is what is pleasant to the virtuous person. It is these pleasures that form part of eudaimonia. People who are not virtuous may get pleasure from other activities, but such pleasure is not good or ‘truly’ pleasant.

# Eudaimonia and philosophy

Aristotle divides reason into practical reason and theoretical reason. We discuss the place and role of practical reason in eudaimonia in the section ‘Aristotle on practical wisdom’. But we have said nothing about theoretical reason. What part does this play in eudaimonia?

Before we turn to the role of theoretical reason, it is worth listing some central claims about eudaimonia.

1. Eudaimonia is not a state, but an activity. You don’t live the best life by being asleep or suffering such misfortune that you can do very little.
2. It is desirable for its own sake and it is self-sufficient.
3. It involves virtuous actions, as these are desirable for their own sake.

Aristotle has also just argued that eudaimonia involves pleasure. But we shouldn’t make the mistake of thinking that the best life is one of pleasant amusements, even if this is what people with power and wealth spend time doing. People find different activities pleasant depending on their character. What is truly pleasant is what is pleasant to the good person, and this is a life of virtuous activity, not a life of mere amusement.

So, to theoretical reason. Theoretical reason – the contemplation of truth – is what is ‘highest’ about human beings, Aristotle argues. Animals have a form of practical wisdom, in that they consider and act on what is best for themselves. But they do not contemplate general truths. This ability is our share in ‘divinity’. Eudaimonia, therefore, must include excellent activity of theoretical reason, which is philosophy.

P1. This activity is best, because theoretical reason is the best thing in us and with it, we contemplate what is best (the greatest, most wonderful and most divine things in the universe), not merely what is best for us (as in practical wisdom).

P2. We are able to undertake this activity more continuously than any other activity, so it leads to the most continuously happy life.

P3. It is the most pleasant activity – at least, its pleasures are most pure and enduring, unlike pleasures of the body.

P4. It is the most self-sufficient activity. Nothing further arises from it (it is knowledge for its own sake), while in other virtuous activities, we normally gain something (honour, gratitude, friendship, power, etc.) beyond doing the action. We need fewer external goods for this than for any other virtuous activity. (To be generous, you need money. To be courageous, you need power. To be temperate, you need opportunities ….)

P5. We are active in order to have leisure. ‘Leisure’ is undertaking those activities we wish to undertake. The virtues of politics aim at creating space for leisure, just as we only undertake war in order to achieve peace. They serve the activity of reason

P6. Finally, theoretical reason is what we most are, it is our characteristic activity.

C1. Therefore, the best and most pleasant life for us, given our nature, will be a life of reason. The life of the philosopher (or more generally, a life dedicated to knowledge) will be the best life.

Aristotle concludes that we should strive to live such a life of theoretical reasoning as far as possible, to live in accordance with the best thing in us. But we are human, and require more than this. Hence the life of virtue more broadly is also part of eudaimonia, as he has argued all along. Having passions, having a body, living with others – these are all characteristically human too. Furthermore, the life of virtue doesn’t require a great deal of external goods, and so while these are necessary, they are not central.

Aristotle on Virtue

As we have seen, Aristotle defines ‘Eudaimonia’ as the good for a human life. It is often translated as ‘happiness’ but Aristotle says it is ‘living well and faring well’. Eudaimonia is ‘the good’ or the ‘good life’ for human beings as the particular sort of being we are. To achieve it is to live as best a human being can live. But what sort of life is a good or flourishing life for us? Aristotle argues that a human life is distinctively the life of a being that can be guided by reason. Qualities that enable us to lead such a life are virtues. In this section, we consider Aristotle’s theory of the virtues.

# The rational ‘soul’

If the good life for human beings is living in accordance with reason, and this requires the virtues, what are the virtues?

A virtue is a trait of a person’s ‘soul’ – we would perhaps say ‘mind’ or ‘self’. Aristotle provides an analysis of the soul. We can divide it into an arational part, and a rational part (at least in analysis, even if there aren’t literal ‘parts’). The arational part can be further divided in two – the part that is related to ‘growth and nutrition’ (Aristotle thought that all life has soul) and the part related to desire and emotion. The desiring part we share with other animals, but in us, it can be responsive to reason. For instance, suppose someone wants to use all their money to buy things they want, but they recognise that it is good to share their wealth with others, and so they do so, their desire gives way. Someone with the virtue of generosity has reshaped their desires, and is not even tempted to try to spend their money on themselves, but happily provides for other people’s needs and desires. What they want ‘speaks with the same voice’ as their reason.

We can talk about the rational part of the soul having two parts as well. There is, again, the desiring part which can respond to reasons and there is the part with which we reason, which has reason ‘in itself’.

Virtues are traits that enable us to live in accordance with reason. They are, therefore, of two kinds – virtues of the intellect (traits of the reasoning part) and virtues of character (traits of the part characterised by desire and emotion). In the rest of this section, we concentrate on virtues of character.

# Virtues as character traits

Aristotle says that anything that is part of the soul (the mind) is either a passion, a faculty or a state (trait) of character. So since virtues are part of the soul, they must be one of these.

1. Passions: Aristotle’s term ‘passions’ covers our bodily appetites (for food, drink, sex, etc.), our emotions, and any feelings accompanied by pleasure or pain. But these can’t be virtues for three reasons.
   1. Just having a particular passion – feeling hungry or angry – doesn’t make you a good or bad person.
   2. We don’t choose our passions, but virtues are related to the choices we make. We cannot generally, just by an act of will, choose what we feel or want.
   3. Virtues concern how we are disposed to feel and act; they are not desires that actually motivate us.
2. Faculties: faculties are things like sight or the ability to feel fear. Virtues can’t be these, since we have these naturally but we have to acquire virtue.
3. So virtues must be states of character.

Aristotle defines states of character as ‘the things in virtue of which we stand well or badly with reference to the passions’. Character involves a person’s dispositions that relate to what, in different circumstances, they feel, how they think, how they react, the sorts of choices they make, and the actions they perform. So someone is short-tempered if they are disposed to feel angry quickly and often; quick-witted if they can think on their feet; intemperate if they get drunk often and excessively. (‘Temperance’ is the virtue relating to pleasure, especially our desires for food, drink, and sex.) What we find pleasant also reveals our character.

Character has a certain stability and longevity. Character traits last much longer and change less easily than many ‘states of mind’, such as moods and desires. But character can change, and so it is less stable and long-lived than personal identity. Yet it is central to being the person one is. Annas comments that my virtues are dispositions of me, and so they are connected to my life as a whole. Aristotelian virtue theory assumes that there is a sense in which I can think about my life as a ‘unity’.

What kind of state of character is a virtue? Some traits of character, such as being short-tempered or greedy, stop us from leading a good life – these are vices. Other traits of character, such as being kind or courageous, help us to lead a good life – and these are the virtues. Any virtue makes the thing which has it good and able to perform its characteristic activity well. So, in us, a virtue of character is a disposition to feel, desire and choose ‘well’, which is necessary if we are to live well and so achieve eudaimonia.

This last point – choosing well – is important. Annas notes that virtues in the fullest sense aren’t simply dispositions to behave in certain ways, like being clumsy is a disposition to have accidents. They are expressed in the choices we make and the reasons for which we act, and a virtue involves a commitment by the person to an ethical value. And so practical reason is central to having and exercising the virtues.

# Virtues, the doctrine of the mean and the importance of feelings

What is involved in choosing and living well? Aristotle compares living well with other activities, such as eating well or physical training. In these cases, the good nutritionist or good trainer needs to avoid prescribing too much food or exercise or too little. We achieve health and physical fitness by following an ‘intermediate’ course of action, which Aristotle calls the ‘mean’. However, what this is differs from person to person. A professional sportsman needs more food and exercise than most people. The mean, what is neither too much nor too little, is relative to each individual. (This ‘mean’ is not a mathematical quantity, an ‘objective’ mean halfway between the two extremes, as 6 is halfway between 2 and 10.)

Now, in the ‘art of living’, so to speak, something similar applies. We can feel our passions either ‘too much’ or ‘too little’ – and here we see the importance of feelings in virtue. Virtue involves being disposed to feeling in an ‘intermediate’ way, neither too much nor too little. Some people feel angry too often, over too many things (perhaps they take a critical comment as an insult), or maybe whenever they get angry, they get very angry, even at minor things. Other people feel angry not often enough (perhaps they don’t understand how people take advantage of them). To be virtuous is ‘to feel [passions] at the right times, with reference to the right objects, towards the right people, with the right motive, and in the right way’. This is Aristotle’s ‘doctrine of the mean’.

It is important to note that Aristotle’s doctrine of the mean does not claim that when we get angry, we should only ever be ‘moderately’ angry. We should be as angry as the situation demands, which can be very angry or only slightly irritated. Given the very close connection between what we feel and how we choose to act, virtues are dispositions of choice as well, and there is a ‘mean’ for actions as well as for feelings.

Julia Annas’ development

Annas expands on this. We need to do the right thing for the right reasons and in an appropriate way. This appropriate way involves both affective (relating to moods, feelings and attitudes) and intellectual aspects. Our action needs to be accompanied by the right feelings. This isn’t merely a matter of self-control, but whole-heartedness. As in our example of generosity above, doing the right thing but grudgingly, or while controlling one’s temptation to do the wrong, while better than doing the wrong thing (!), is not virtuous, according to Aristotle.

The intellectual aspect of virtuous action involves understanding that this action is the right thing to do. What the right action, time, object, person and so on is, for both feeling and action, Practical wisdom helps us to know. (We won’t complete our account of virtue, therefore, until we have understood what practical wisdom is.) Practical wisdom is a virtue of reason. Our passions, we noted, are susceptible to reason. There can be right and wrong ways to feel passions, and the right way to feel passions is determined by reason. If we feel our passions ‘irrationally’ – at the wrong times, towards the wrong objects, etc. – then we don’t live well. Likewise, we can choose the right or wrong actions and act for the right or wrong reasons, usually as a result of whether the feelings that help influence our choices are themselves rational or irrational. So, Aristotle concludes, a virtue is ‘a state of character concerned with choice, lying in the mean, i.e. the mean relative to us, this being determined by a rational principle, and by that principle by which the person of practical wisdom would determine it’.

## The application of the doctrine of the mean to particular virtues

The doctrine of the mean entails that we can (often, if not always) place a virtue ‘between’ two vices. Just as there is a right time, object, person, etc., at which to feel fear (or any emotion), some people can feel fear too often, about too many things, and towards too many people, or they get too afraid of things that aren’t that dangerous. Other people can feel afraid not often enough, regarding too few objects and people. Someone who feels fear ‘too much’ is cowardly. Someone who feels fear ‘too little’ is rash. Someone who has the virtue relating to fear is courageous. The virtue is the ‘intermediate’ state between the two vices of ‘too much’ and ‘too little’.

Aristotle presents the following examples. For many states of character, he notes, we don’t have a common name.

|  |  |  |  |
| --- | --- | --- | --- |
| Passion/concern | Vice of deficiency | Virtue | Vice of excess |
| Fear | Cowardly | Courageous | Rash |
| Pleasure/pain | ‘Insensible’ | Temperate | Self-indulgent |
| Giving/taking money | Mean | Liberal (‘free’) | Prodigal (‘spendthrift’) |
| Spending large sums of money | Niggardly | ‘Magnificent’ | Tasteless |
| Important honour | Unduly humble | Properly proud | Vain |
| Small honours | ‘Unambitious’ | ‘Properly ambitious’ | ‘Overambitious’ |
| Anger | ‘Unirascible’ | Good-tempered | Short-tempered |
| Truthfulness (regarding oneself) | Falsely modest | Truthful | Boastful |
| Humour | Boorish | Witty | Buffoonish |
| Pleasant to others | Quarrelsome, surly | Friendly | Obsequious |
| Shame | Shy | Modest | Shameless |
| Attitude to others’ fortune | Spiteful (rejoicing in others’ bad fortune) | Righteously indignant (pained by others’ undeserved good fortune) | Envious (pained by others’ good fortune) |

Obviously, Aristotle notes, not all types of actions or states of character can pick out a mean. For example, being shameless is not a mean, but a vice, while murder is always wrong. Furthermore, we often oppose a virtue to one of the two vices, either because it forms a stronger contrast with that vice (e.g. courage–cowardice) or because we have a natural tendency towards that vice, so need to try harder to resist it (e.g. temperance–self-indulgence).

# The role of education in the development of a moral character

We now know what virtues are. But how do we acquire them? Virtues are necessary for eudaimonia, but because they are dispositions towards feeling passions, and passions are not under the direct control of the will, we can’t simply choose to become virtuous. Aristotle argues that we acquire virtues of character through ‘habit’, in particular, the habits we form during our upbringing. We need to develop virtue because, Aristotle argues, we are not virtuous just by nature. He points out that for what we can do naturally, we first have the ‘potentiality’ and then exhibit the activity. For example, you don’t acquire sight by seeing; first you have sight, then you can see. But for the virtues, you must first practice acting in a virtuous way – courageously, generously, kindly, etc. – before you can be virtuous. We are not naturally virtuous, but we are naturally capable of becoming virtuous.

## The skill analogy

We can understand how we acquire virtues by an analogy with acquiring practical skills, such as carpentry, cookery or playing a musical instrument. There are two parts to the analogy.

The first part regards how the development of the skill/virtue begins. We come to form dispositions to feel and behave in certain ways by what we do. The same is true for practical skills. You cannot learn cookery or a musical instrument just by studying the theory, by merely acquiring knowledge about how to cook or play; you have to practice the activity. Likewise, being told how to be good is not enough to become good; you have to actually practice being good: ‘the virtues we get by first exercising them, as also happens in the case of the arts as well [e.g. learning to play a musical instrument]. For the things we have to learn before we can do them, we learn by doing them.’ Hence, ‘by doing the acts that we do in our transactions with other men we become just or unjust, and by doing the acts that we do in the presence of danger, and by being habituated to feel fear or confidence, we become brave or cowardly’.

It can seem that there is a puzzle in what Aristotle says. In order to become just, we have to do just acts. But how can we do just acts unless we are already just? The puzzle is solved by distinguishing between actions which are ‘in accordance with’ justice and just acts, properly so called. The actions that we do when learning to become just are acts in accordance with justice. But a just act is an act that is not only in accordance with justice, but also done as the just person does it.

This takes us to the second part of the skill analogy, which Annas draws out. While we first learn from others, this is only half the process. As with practical skills, the aim in moral education is to get the child/student to learn to think for themselves. This involves two related skills.

First, the expert progresses from simply following rules to developing a highly attuned sensitivity to how each situation is different, and how to respond to those differences appropriately. So a good carpenter responds and works with the knots and grain in the wood; a good chef checks the seasoning of each dish by tasting it, and adjusts each element of a meal in light of the others. Likewise, in growing in virtue, we become better able to recognise situations in which action is called for and what to do in response.

Second, just as in developing a practical skill, we understand why this way rather than that way is better – why this screwdriver, why this spice, and not that one – so, in its development of virtue, the child comes to reflect on the reasons for acting this way rather than that. He or she tries to make their moral judgments and practice more coherent and unified and is able to justify their choices. In all this, our appreciation of what is virtuous may change. For example, we may first associate courage just with physical courage, particularly in fighting. But then, with reflection and greater experience, we come to identify it in dealing with emotional challenges, with loss, in friendships and speaking truth to power. We may come to rethink whether certain types of action that we thought exemplified courage, e.g. in Hollywood blockbusters, really are courageous at all or a form of masculine competition. Thus as we develop in virtue, we understand, in a practical way, more about what is good. All this is part of our developing practical wisdom.

And so, Aristotle argues, a fully virtuous action is one in which the agent knows what they are doing, chooses the act for its own sake (i.e. for the end at which the relevant virtue aims, e.g. justice), and makes their choice from a firm and unchangeable character. Until it has moved from the early learning stage to the development of a richer expertise, a child may do what is just (such as not taking more than its fair share) because it is told to do so; or because it likes the person it is sharing with; or because it wants to please an adult, and so on. It neither truly understands what justice is nor does it choose the act because the act is just.

Annas notes that there are limits to the skill analogy. First, we don’t have to pursue the ends of a particular skill, such as creating beautiful furniture or tasty food, e.g. we may lose interest in that activity or its rewards. But virtues pursue our final end – eudaimonia – and we can’t opt out of having that end, we can’t want to live at all and cease to want to live a good life. Second, many practical skills can be developed without involving our emotions, but the development of our emotional dispositions is central to developing virtue.

Aristotle concludes that whether or not we can lead a good life depends a great deal on the habits we form when we are young – in our childhood and early adulthood. Furthermore, because our character is revealed by what we take pleasure in, we need to learn to take pleasure in the things that we should take pleasure in, and be pained by what should pain us.

Aristotle on voluntary action, choice and moral responsibility

The ideas of voluntary action, choice and responsibility are important in our moral lives. As Aristotle notes in the *Nicomachean Ethics*, we praise and blame what is voluntary, but not what is involuntary. But what is the difference between what is voluntary and what is not, and what is it to be responsible for one’s choices and actions?

# Voluntary and involuntary actions

There are, says Aristotle, two things that render our actions involuntary – force and ignorance. When we act voluntarily, by contrast, we know what we are doing, and we bring it about ourselves. Contrast three cases of standing on a train and stepping on someone’s foot:

1. The train lurches, you lose your balance, and accidentally step on someone’s foot. Stepping on their foot is involuntary, caused by force.
2. You shuffle your feet to get comfortable, and put your foot down on someone’s foot without looking. Although moving your feet is voluntary, stepping on someone’s foot is involuntary, caused by ignorance (that their foot was there).
3. You deliberately and knowingly bring your foot down on top of someone else’s. This is voluntary.

## Force

We can be forced to do things not only by physical forces but also by psychological pressure (such as threat of pain). Where no one could withstand such pressure, we don’t blame someone for what they do. This shows that what they do is involuntary. However, the psychological pressure must be negative not positive. We don’t think of the prospect of something good or pleasant as ‘forcing’ us to act. Giving in to temptation is not acting involuntarily! When we act involuntarily, we do so with pain and regret.

Now, some actions that we do, we don’t want to do. These might be called voluntary or involuntary. Aristotle gives the example of sailors throwing goods overboard in a storm. They want to save the boat, but they don’t want to lose the goods. Such actions, he argues, should be called voluntary. First, actions which we do to avoid a greater evil or in order to secure some good end are the right actions to choose. Second, we praise people for such actions, and we noted above that praise and blame attaches to what is voluntary.

So, the distinction between voluntary and involuntary actions relates to the moment of action in the particular circumstances one is in, not whether the action is generally desirable.

## Ignorance

Some actions done as a result of ignorance are involuntary, some are simply ‘not voluntary’. The difference lies in whether the action is one that causes us pain or regret. Suppose, again, you step on someone’s foot while shuffling your feet. If you regret this, then stepping on their foot is involuntary. But if you don’t care, then it is simply non-voluntary.

The kind of ignorance that makes an act involuntary relates to the particular circumstances of the action. You know what you are aiming at (you aren’t ignorant of the end, e.g. ‘to get comfortable’), and you can know relevant general truths (e.g. people have feet). But you don’t know the particular circumstances of the action; e.g. what you are actually doing (stepping on someone’s foot), or in other cases, what its consequences will be, what tools you are using to act with, or how (in what manner) you are acting (e.g. you might think you are gently helping, when you are actually annoyingly hindering).

We should also distinguish acting in ignorance from acting as a result of ignorance. When drunk or really angry, you may do something without fully understanding just what you are doing. Here we say that your action is a result of your drunkenness or rage, rather than your ignorance, and we may still blame you for your actions. But your drunkenness or rage puts you in a state of ignorance. So you act in ignorance, but not from ignorance.

## Voluntary action

Voluntary action, then, is action that you bring about, in the knowledge of what you are doing.

Sometimes people say that actions done from desire or emotion aren’t voluntary. But this is a mistake for four reasons.

1. If it were true, we would have to say that neither animals nor children ever act voluntarily.
2. There are many good actions that we can do from desire and emotion (such as being kind), and we ought to do them. It would be strange to say that what we ought to do is not voluntary.
3. Actions done from desire or emotion are pleasant, not painful. But we said involuntary actions are painful, while the prospect of what is pleasant does not force us.
4. Our desires and emotions are no less part of us than our reason. Acting on them is something we do.

# Choice and deliberation

We need to distinguish what is voluntary from what we choose. Everything we choose to do is voluntary, but not everything voluntary is chosen. For instance, spontaneous actions and the actions of young children and animals are voluntary, but not chosen in the sense intended here. So what is choice?

1. It isn’t desire – someone who gives in to temptation acts with desire, but not from choice, while someone who resists temptation acts on choice, but against their desire.
2. It isn’t ‘wish’, since you can wish for what is impossible and things you can do nothing about, but choice relates to what we can actually do. What we wish for is also an end, something we are aiming at. What we choose are not ends, but the means to achieve our ends.
3. It isn’t a kind of opinion – opinions are true or false, but choices are good or bad.
4. Instead, choice relates to voluntary action, where this is done on the basis of deliberation.

So what is deliberation? We don’t deliberate about what we can’t change, such as the facts – we investigate these (theoretical reasoning). We only deliberate about things that we can change (practical reasoning). In fact, we only deliberate when we need to act differently on different occasions. You don’t deliberate about how to make a cup of tea (once you’ve learned) – you just get on and do it! So deliberation is a kind of reasoned thought about what we can change by our efforts, and where we need to act differently on different occasions.

Aristotle also claims that we don’t deliberate about ends. But is this right? For instance, I might study in order to get a good grade (my end). But I might well deliberate about whether to get good grades; for example whether it is worth the effort. Or again, I may have two ends that conflict – being a good friend and telling the truth – and I deliberate about which end to pursue.

However, what Aristotle probably means is that we don’t deliberate about ends as ends. When we deliberate, we always have some end in view, and whatever we are considering is as a means to that end. If I deliberate about whether to get good grades, I am considering this in light of some further end, such as going to university. If I deliberate about being a good friend or telling the truth, I do so in light of my final end – leading a good life.

We can now say what choice is. Choice is what we decide upon as a result of deliberation. So it is a deliberate desire regarding something that is in one’s power.

# Moral responsibility

When are we morally responsible for what we do? This question is important to ethics. We are generally happy to say that people who do good actions are morally responsible for what they do, and we praise them accordingly. But do people who are bad do bad things voluntarily and by choice? Before Aristotle, Socrates had argued that they do not. Everyone aims at what they believe is good. All bad action is acting from ignorance of what is truly good, so it is not voluntary. Aristotle accepts that bad people are ignorant of the good, but maintains that they still act voluntarily.

To know fully what the right act is involves understanding why it is right. Someone who is bad might know, as a child does, that action x shouldn’t be done. But if they don’t understand why, they don’t really know what they ought to do. Put another way: given that we all aim at eudaimonia, what is good is the ‘proper’ object of wish – what is truly desirable (§4). This is, in fact, what the good person desires. Bad people desire what is not truly desirable, but they don’t realise this. Most errors of this kind are caused by pleasure. What is bad can seem desirable if we think it is pleasant. And different states of character find different things pleasant; e.g. the just person finds justice pleasant, but the unjust man does not.

However, the fact that bad people are ignorant of what is good does not entail that bad people act involuntarily. Aristotle offers four arguments for this claim.

1. We noted that choice relates to the means, the actions that we take. What it is in our power to do, it is also in our power not to do. So we can choose to do either good or bad actions. So bad people do bad actions voluntarily.
2. We encourage people not to do bad actions, yet we don’t encourage people not to do things that are out of their power. That would be pointless. So bad actions are done voluntarily.

To these arguments, we might respond that there is a sense in which bad people choose to do bad actions. But still they are not morally responsible for them, because they are pursuing what seems good to them. They do not know what is truly desirable, and it is this ignorance that influences their choices.

Aristotle’s third argument responds to this objection.

1. Bad people became bad as a result of their choices. Therefore, they are responsible for becoming bad, and thus becoming ignorant of what is good.

Why believe this? Because we acquire a particular state of character by acting in a corresponding way. For example, we become just by acting in accordance with justice. Thus, we are partly responsible for our character traits. We can choose how to act, knowing that how we act will make us good or bad people. A person, through choosing to act badly, becomes a bad person, and at that point, they have become ignorant of what is good.

Rather like becoming drunk and then not knowing what you are doing; or becoming ill through ignoring medical advice; or becoming ugly through lack of care and exercise; we are responsible for becoming bad through the choices we made. We can’t, when drunk, choose to be sober; or when ill, choose to be healthy; or when ugly, choose to be beautiful; so when bad, we can’t simply choose to become good. Yet despite this, our condition is voluntary and we are morally responsible for it. What appears good or pleasant depends on one’s character traits. If the bad person is mistaken about what is good, this is as a result of their character traits. But as they are responsible for their character traits, they are responsible for their lack of knowledge of what is truly good. So the fact that they are doing something bad, thinking that it is good, does not count as the kind of ignorance involved in involuntary action (acting from ignorance), but as the kind of blameworthy ignorance (acting in ignorance) involved in drunkenness.

1. If we reject this argument, and claim that the bad person is not responsible for what they think is good, then we must apply the claim generally – no one is responsible for what seems good or bad to them.

P1. If the bad person is not responsible for their bad actions, and these are not done voluntarily, then the good person is not responsible for their good actions, and these are not done voluntarily.

P2. But we said earlier that what is good cannot force us to act, and that what is involuntary is painful and causes regret.

C1. So good actions are done voluntarily.

C2. Therefore, so are bad actions.

Of course, actions and character traits are not voluntary in the same way. Voluntary actions are under our control from start to finish. But with the development of character traits, it is only at the beginning – in choosing the actions that lead to certain character traits – that they are fully voluntary. After this, we gradually become a certain sort of person, and then we cannot simply choose to be a different sort of person.

# Practical wisdom

Practical wisdom (phronesis) is an intellectual virtue, a virtue of practical reasoning. Aristotle draws a distinction between theoretical reason and practical reason. Roughly, theoretical reason investigates what we can’t change and aims at the truth. Practical reason investigates what we can change and aims at making good choices. Reasoning about what we can change is deliberation, so practical reason is expressed in deliberation. To make good choices, not only must our reasoning be correct, but we must also have the right desires.

The person with practical wisdom deliberates well about how to live a good life. So practical wisdom is ‘a true and reasoned state or capacity to act with regard to the things that are good or bad for man’.

## The role of practical wisdom

Practical wisdom differs from other sorts of knowledge both because of its complexity and its practical nature. Aristotle claims that it involves

1. a general conception of what is good or bad, related to the conditions for human flourishing;
2. the ability to perceive, in light of that general conception, what is required in terms of feeling, choice and action in a particular situation;
3. the ability to deliberate well; and
4. the ability to act on that deliberation.

So it involves general knowledge, particular knowledge, an ability to reason towards a choice, and an ability to act on that choice.

There are different ways in which we can fail to deliberate well.

1. We can deliberate with the wrong end. Our starting point is wrong, and so our choice is wrong. Our general knowledge of the good is faulty.
2. We can have the right end, and perhaps even achieve it. However, we don’t understand the right means to the end, and so if we achieve our end, this is accidental or lucky. Either our knowledge of the particular circumstances or our reasoning is faulty.
3. We can fail to deliberate when we should or take too long.

So practical wisdom means deliberating with a good end, identifying the right means, and doing so in a timely way. In its fullest sense, practical wisdom involves deliberating from the most unqualified end, eudaimonia itself.

## Practical wisdom and rules

Point (2) above says that practical wisdom involves understanding what is required in a particular situation in light of a general understanding of what is good. The question that faces us on any occasion is how to achieve what is good – what the good life involves – in the here and now, in this situation. But there are no rules for applying knowledge of the good life to the current situation. What is right on a particular occasion is in accordance with ‘right reason’, but Aristotle has argued that this can vary from one occasion to another. Furthermore, this kind of insight is inseparable from making a good decision: we must not only understand the situation (which can involve considerable sensitivity), but also understand how to act well in it.

This makes it impossible to make generalisations about right and wrong, good and bad, that are true in all cases. Practical wisdom intuitively ‘grasps’ the particular facts involved in the case. This does not make ethics subjective, as there is a truth of the matter to be discovered. However, proving the truth of one view against another is not possible by argument alone. If you are blind, I may not be able to convince you of the colour of moonlight; if you lack insight into what is good, I may not be able to convince you of the goodness of being kind. If you can’t understand the situation we are facing, I may not be able to convince you that the right thing to do on this occasion is to be generous.

And so, Aristotle argues, practical wisdom is not something that can be taught, for what can be taught is general, not particular. Rules and principles will rarely apply in any clear way to real situations. Instead, moral knowledge is only acquired through experience.

# The relation between practical wisdom, virtue and action

How does practical wisdom relate to virtue and virtuous action? We can imagine this objection: living a good life is a matter of being good, and this involves the virtues. So what use is practical wisdom?

A first, simple answer is this: the virtues (justice, courage, generosity, etc.) set our ends. Because we are virtuous, we aim at the good life, and we have a reliable conception of what this is (it involves justice, courage, generosity, etc.). But that isn’t enough to live a good life, because it doesn’t tell us what is good (courageous, etc.) in this particular situation. For that, we need practical wisdom to identify the (constitutive) means to our virtuous ends.

Further reflection tells us more. We can draw a distinction between acting in accordance with a virtue – roughly ‘doing the right thing’ – and doing a fully virtuous action. A fully virtuous action is one in which the agent knows what they are doing and chooses the act for its own sake. Both this knowledge and this kind of choice depend on having practical wisdom. The knowledge involves understanding what is good in this situation, and choice depends upon deliberation, and good deliberation involves practical wisdom. So acting virtuously requires practical wisdom.

So Aristotle draws a distinction between ‘natural’ virtue and ‘full’ virtue. He allows that we can have good dispositions from birth; e.g. someone might be naturally kind. But this doesn’t amount to ‘full virtue’. A naturally kind child doesn’t fully comprehend the nature of their action, and could easily be misled into being kind for the wrong reasons or at the wrong time. Without practical wisdom, we can’t have full virtue.

But practical wisdom also depends on virtue. It is possible to deliberate from the wrong ends. A bad person can be very clever in achieving what they want. But cleverness is not practical wisdom, because practical wisdom also involves having general knowledge about what is good. This depends upon being virtuous, because what appears good to someone depends on their character traits. So on Aristotle’s theory, we become both good and practically wise together.

# Objection 1: Guidance on how to act

A first issue facing Aristotle’s virtue ethics is whether it can provide us with any helpful guidance on how to act. What does Aristotle say that can help us decide what to do?

Many philosophers have thought that Aristotle’s doctrine of the mean should function in this way. But it isn’t much help. First, ‘too much’ and ‘too little’ aren’t quantities on a single scale. The list of ‘right time, right object, right person, right motive, right way’ shows that things are much more complicated than that. Second, it gives us no help with understanding, for example, how often we should get angry, and how angry we should get. Just about anything could be ‘in the mean’ if the circumstances were right!

But it is unlikely that Aristotle intended the doctrine of the mean to be helpful in this way. We can’t ‘figure out’ what it is right to do by applying a rule like the doctrine of the mean; we must have practical wisdom. Aristotle says explicitly that what is in the mean is ‘determined by the person of practical wisdom’. And life is complicated; so practical wisdom isn’t about applying easy rules either. It’s about ‘seeing’ what to do, which requires virtues of character and lots of experience.

But does Aristotle’s theory of practical wisdom provide any guidance about what to do? If I have practical wisdom, it seems that I simply know what to do. But if I do not have practical wisdom, what then? Knowing that the right action is what a virtuous person would do doesn’t help me, because I don’t know what the virtuous person would do! Aristotle seems to admit as much when he says that practical wisdom requires virtue. Without a good character, I cannot understand what is truly good. But this means that knowledge of the good is not within everyone’s reach. Either Aristotle’s theory provides no guidance to anyone who isn’t virtuous, or his theory is wrong because we are all sufficiently rational to understand what is right and wrong.

Aristotle argues that this is too simple. Knowledge of the good can come in degrees, and we can improve or destroy our ability to know what is good by the kind of character we develop. If someone has a completely depraved character, perhaps they really don’t know what is good or bad. But most people will have enough understanding of the good to make moral decisions. Furthermore, people can improve their knowledge of what is good by becoming more virtuous people.

In her article ‘Virtue ethics’, Julia Annas argues that virtue ethics assumes that each of us already has a life by the time we start to reflect on which action is the right one. This has two implications. First, we already have some general guidance from the culture in which we grow up, but reflection will reveal that our traditional ethical views are inadequate in some way or other. Our desire to do what is right is an expression of our striving to be better people. Second, we are each at different stages in ethical development and have different aims and ideals in life. Reflecting on what to do, therefore, can’t be like the matter of following an algorithm, like learning to use a computer. There simply cannot be a specification of the ‘right action’ that is universal, the same for everyone and available to everyone, regardless of what they are like as people already. Becoming virtuous takes experience and practice. No teacher or book can make you virtuous (‘follow these simple rules, and you will achieve eudaimonia’).

The objection that virtue ethics can’t provide guidance on how to act is thinking of guidance too much in terms of rules. Just because practical wisdom is not a set of rules, that doesn’t mean it provides no guidance at all. Aristotle’s theory suggests we think about situations in terms of the virtues. We can ask a series of questions: ‘would this action be kind/courageous/loyal … ?’ or again, since I am deciding how I should act, not how anyone should act, ‘what would I do if I were more kind/courageous/loyal…?’ Thinking about what to do in this way could be very helpful.

# Objection 2: Conflicts between virtues

A second issue for Aristotle’s virtue ethics regards cases of conflict between virtues. For example, can we show justice and mercy, or do we have to choose?

Aristotle denies that conflicts between virtues ever take place. You need practical wisdom to understand what each virtue actually requires you to do in this particular situation. With such understanding, you will be able to discover a path of action which satisfies the demands of each virtue that is relevant to the situation. If you think that mercy requires injustice, or that justice demands being merciless, then you have misunderstood what justice or mercy actually mean in this situation. For example, perhaps we are motivated towards mercy in rectifying an injustice when someone appeals to difficult circumstances or ignorance of the effects of what they did. On Aristotle’s analysis, such factors are directly relevant to judging the injustice of the act (whether it is unjust, or done unjustly, or done by an unjust person). So they are relevant to what justice requires of us.

Aristotle explicitly rejects the claim that morality involves absolute or universal rules. It is all a matter of context and judgement, and the idea that we are always pursuing the final end of eudaimonia provides a framework in which to make such judgments. Virtues don’t make demands of their own accord, but provide us with various means to achieve eudaimonia. All this makes it easier to resolve potential conflicts.

Nevertheless, whether the theory is convincing in all cases can only be judged by looking at possible counterexamples. For example, could loyalty to a friend ever require you to be dishonest?

# objection 3: The possibility of circularity involved in defining virtuous acts and virtuous people in terms of each other

A third issue relates to Aristotle’s accounts of virtuous action and the virtuous person. A simple reading, which causes the problem, is this:

1. an act is virtuous if it is an act that would be done by a virtuous person in this situation;
2. a virtuous person is a person who is disposed to do virtuous acts.

The difficulty with these definitions is that, taken together, they do nothing to clarify what a virtuous act is or what a virtuous person is. For instance, if we substitute the definition of a virtuous person in (1), we get ‘an act is virtuous if it is an act that would be done by a person who is disposed to do virtuous acts in this situation’. The definition is circular, because we have used the term ‘virtuous act’ to define what a virtuous act is! We get the same problem if we substitute the definition of a virtuous act in (2): ‘a virtuous person is a person who is disposed to do acts that would be done by a virtuous person’.

One way to solve the problem is to pay closer attention to Aristotle’s definitions. A (fully) virtuous act is indeed an act that a virtuous person does, when they know what they are doing and choose the act for its own sake. However, a virtuous person is not simply someone who does virtuous actions. A virtuous person has the virtues, which are traits, including states of character and excellences of reason, that enable them to achieve eudaimonia. States of character relate to our choices and actions, but they are equally concerned with our passions and with what we find pleasure in. And eudaimonia is defined not in terms of virtuous actions, but in terms of many activities ‘of the soul’, including feeling, thinking and choosing. So while (1) is correct, (2) is too simple.

We could press the objection a different way. We can’t tell whether an act is virtuous without knowing whether a virtuous person would do it. And we can’t tell whether someone is virtuous without seeing whether they do virtuous acts.

In reply, first, it is true that the criterion for an act being virtuous is that it is an act that a virtuous person would do. But we have a good idea of what a virtuous person is without being able to name particular individuals as virtuous or not. When considering ‘what the virtuous person would do’, we need not have any specific virtuous person in mind. So to judge whether an act is virtuous, we don’t need to first judge that person A is virtuous and then figure out what A would do.

Second, it is true that we infer that someone is virtuous from what they do. But again, this is not the only evidence we have. Virtue is also expressed in emotional responses and pleasure, as well as the quality of someone’s thinking. So there is no circularity in establishing whether an act or a person is virtuous.

Annas provides a different, and much simpler, line of response. Virtue ethics does provide an account of virtuous action by appealing to what the virtuous person would do. But this shouldn’t be understood (as it is in (1) above) as a definition of ‘virtuous action’ that uses some independent definition of ‘virtuous person’. Instead, as we saw above, the account encourages us to think of what I should do in my situation (virtuous action) in terms of what I would do if I were more virtuous (virtuous person). Instead of thinking of the matter from the ‘third person’ (‘what is the definition?’, ‘how can we tell a virtuous action?’), we should think of it from the first person (‘what should I do?’).

OBJECTION 4: VIRTUE AND EUDAIMONIA

In this section, we discuss Aristotle’s theory of the relationship between eudaimonia and virtue. Aristotle argues that the good life for human beings is eudaimonia. In developing his theory, he doesn’t draw a distinction between a life that is good for me and a life that is morally good. But we do commonly draw a distinction between what is in our self-interest and what is morally required, and many people feel that the two can come into conflict. What is the relationship between my living a good life and moral virtue?

Morality can require self-sacrifice; sometimes the morally right thing to do – and so the thing that the virtuous person would do – requires me to give up something good for myself and perhaps even harm myself. For example, giving to charity is important, but how does it contribute to a good life for me? Or again, if someone lives under an unjust dictatorship, courage and justice may lead them to stand up for what is right but be imprisoned or killed as a result. It seems, then, that eudaimonia and the morally good life can come apart. If so, this is an objection to Aristotle’s theory.

This objection leads to a second. Is Aristotle’s account of virtue, as a trait that contributes to the individual’s eudaimonia, correct? Can there be virtues that are not in our self-interest?

In her book *Virtue Ethics*, Christine Swanton provides some potential counterexamples to Aristotle’s theory of virtue. A woman works as a medic in a foreign country, ceaselessly saving lives and relieving suffering, often far from civilization and under difficult circumstances. She is often ill and tired, and doesn’t experience joy in her work, just the conviction that it is needed. She dies prematurely from a virus. Her life looks like a morally good life, one in which demonstrates many virtues – of kindness, compassion, generosity, perhaps justice – but not one in which she, personally, ‘flourishes’ or ‘lives well’. Another example: A man is dedicated to preserving the environment. He works hard to publicise the dangers of our current way of life, but finds that other people simply won’t listen to him. He becomes very stressed and dies, in despair, of a heart attack. His commitment to a good cause looks virtuous and he practices honesty and justice, yet he hasn’t flourished.

Swanton argues that we shouldn’t require virtues to contribute to the eudaimonia of the person who has them. There are other values (other ‘final ends’), including the good of others, the environment, knowledge, art, and so on. Virtues are dispositions to respond to and pursue these values in appropriate ways. Because sometimes these values call on us to sacrifice our own well-being, Aristotle is wrong to think virtues are traits that contribute to one’s own eudaimonia.

In her article ‘Virtue ethics’, Julia Annas provides an extended response to these objections, arguing that they misunderstand the concept of eudaimonia. Our contemporary concept of ‘self-interest’, like ‘happiness’, is far too narrow. If we read Aristotle as equating eudaimonia and self-interest, then the claim that eudaimonia is our final end is a form of egoism. And philosophers have mistakenly objected that Aristotle makes ethics self-centred, all about the ‘best life for me’. But in Aristotle, we don’t have an account of what eudaimonia – flourishing – is for human beings before considering the virtues. We can’t first specify what it would be to live ‘a good life for me’ and then investigate whether having virtues would enable this life or not. Aristotle’s argument is that living in accordance with the virtues gives us the best account of what counts as a flourishing life. If someone else claims that eudaimonia involves getting what you want, having money and power, then they haven’t agreed with Aristotle about eudaimonia but disagreed about whether the virtues are necessary – they have a different conception of eudaimonia to start with.

To aim at getting what you want without concern for the interests of others is egoism. On Aristotle’s virtue-based account of what it is to flourish, aiming at eudaimonia isn’t egoism. Someone who aims to live in a fair, generous, courageous, just way isn’t self-centred. Having these virtues means being committed to other people’s well-being, and having these commitments is part of what it is to flourish. To live this best life, you must treat people in certain, morally good ways. We can illustrate this with friendship. It is a very important part of leading a good, happy life that one has friends. But someone who is a friend just out of self-interest is not a real friend. He will miss out on the good things – the feelings, the character, the state of mind – that come from being a real friend. On the other hand, someone who does not find friendship a beneficial and important part of their life – who is a friend without feeling that they gain from it – is also missing out on what is important in friendship.

Aristotle’s theory entails that the flourishing life involves commitments to values beyond one’s own interests, narrowly defined. Each person aims at their own flourishing, says Annas, just in the sense that each person leads their own life. It makes no sense to think of your eudaimonia as the final end of my life. Eudaimonia isn’t a state of person that someone else can bring about, but a quality of the activity of living life. Only you can live your life and I can live only my life. But in pursuing (my) eudaimonia, I needn’t privilege my interests above those others, and being virtuous may well lead one to stand up for what is right when it doesn’t ‘benefit’ oneself in any narrow sense.

Does this answer the objections? We have a general account of Aristotle’s answers the questions ‘what is good for me?’ and ‘what is morally good?’ together, so that any supposed example of eudaimonia that doesn’t involve the virtues simply begs the question. But the counterexamples still retain some force – aren’t there virtuous lives that don’t involve the person’s flourishing? One response is that the unhappy medic and environmentalist both live lives of integrity. It would be a mistake to say that their lives would be better for them if they hadn’t acted on the values that they hold most dear.

But, pressing the objection, surely they still lose something good from their lives by living as they do; we still have an idea that their lives could have gone better for them. Integrity is not the same as flourishing. Perhaps what Swanton’s objection shows is that, given the different things that are good, what is really wrong with Aristotle’s notion of eudaimonia is that ‘the good life for human beings’ is not a coherent unitary whole. There is no single, complete final end. Virtues can pull in different directions, and flourishing in one sense may lead to not flourishing in another sense.

2. Applied Ethics

Ethical theories are intended to guide us in knowing and doing what is morally right. It is therefore very useful to consider theories in relation to practical issues, in order to understand the theories and their implications better.

The purpose of this section is to think about how three normative theories – utilitarianism, Kantian deontology and Aristotelian virtue ethics would respond to the following ethical issues:

1. Stealing

2. Eating animals

3. Telling lies

4. Simulated killing

**2.1. Stealing**

To steal is to take someone else’s property with no intention of returning it and without their permission (or without the legal right to do so). To own property involves a system of rights. I own my books, for instance. This means that I am free to use them (read them, keep them where I want, etc.), but no one else is free to use them unless I say so. And so because property involves individual rights, whether or not we may steal is a matter of justice, of what each person is ‘due’. I am ‘owed’ what I own, e.g. it should be returned to me if borrowed, and so to deprive me of my property through stealing or refusing to return it looks like a violation of justice. We start from the presumption that stealing is wrong, but arguments might overturn that presumption. What do our three theories say?

# Utilitarianism

Act utilitarianism, in its simplest form, says that an action is right if it maximises happiness, and wrong if it does not. So it does not phrase the rightness or wrongness of an action in terms of justice. If stealing, on some occasion, creates greater happiness than not stealing, then it is morally right on that occasion. Otherwise, it is morally wrong. It may well be that stealing usually leads to more unhappiness in the world, on balance, so it is usually wrong.

But we also recognise that it can be morally permissible to steal in certain situations. For instance, we might say that it is okay to steal in order to save a life or when we are in dire need, and especially if we steal from someone wealthy and so won’t be significantly harmed by the loss of their property. Or again, we may think that it is okay to steal from the wealthy where property is unjustly distributed, especially if we intend to benefit the poor – a ‘Robin Hood’ kind of stealing. Furthermore, we may argue that it would be right to steal something that the owner intended to use for harm, e.g. stealing a gun from someone planning a murder. Act utilitarianism can argue that it recognises these intuitions – these are all occasions on which stealing would lead to greater happiness than not doing so.

However, a society in which people stealing was permitted whenever it increased happiness could indirectly lead to greater unhappiness. People would become afraid that others will steal from them whenever the thief convinced himself that it would bring about greater happiness. In such a society, people would cease to trust each other, and so overall, the act utilitarian approach to stealing would diminish happiness.

John Stuart Mill recognises the importance of property rights and the importance of rights to feeling secure. Having and respecting property rights contributes most to happiness in the long term, and Mill comments that justice is ‘most sacred and binding part of all morality’. This suggests that we should never violate someone’s rights for some other purpose, e.g. maximising happiness on this occasion, and so we should never steal. But the examples given above, of Robin Hood or stealing in dire need, indicate that such a strict rule is counterintuitive. There need to be exceptions.

How would rule utilitarianism deal with this? Rule utilitarianism says that an action is right if, and only if, it complies with those rules which, if everybody followed them, would lead to the greatest happiness (compared to any other set of rules). Perhaps there should be no property – perhaps the world would be a happier place if no one owned anything? If so, rule utilitarianism would say that we should have a rule forbidding ownership. However, this isn’t relevant to whether *stealing* is morally right. Stealing assumes that there is property – no property, no theft. You can’t steal something if it doesn’t belong to someone else. If we want to bring about a world without property, stealing is not the way to do so – we will need new laws. Meanwhile, while there is property, people become very upset if it is stolen from them. So we cannot justify stealing by arguing that the world would be better off without property. We can only argue that a world in which stealing is impossible would be happier than a world in which there is property. But if there is property, the rule ‘Do not steal’ creates more happiness than a rule that allows stealing freely.

Is there a better rule regarding stealing, one that allows exceptions? Like Mill, rule utilitarians look at the long term consequences, and following a clear, simple rule will probably lead to better consequences overall than trying to build in lots of exceptions. Although the simple rule won’t lead to the best consequences in every case, exception clauses may tempt people to think that their situation is ‘exceptional’ and lead them to steal when they should not – and so a rule with exceptions won’t, in practice, lead to the best consequences in every case either, because people won’t follow it.

One solution for rule utilitarianism is to keep the simple rule against stealing, but add another, ‘general purpose’ rule: ‘in circumstances in which not doing something will lead to significant and immediate harm, then do that thing even if that means breaking some other rule’. This would allow stealing to prevent significant and immediate harm.

# Kantian deontology

Kant argues that we should ‘Act only on that maxim through which you can at the same time will that it should become a universal law’ – his ‘Categorical Imperative’. Acting on the maxim ‘To steal something I want if I can’t afford it’ leads to a contradiction when universalised. Stealing can only be the right thing to do if everyone could do it. However, if we could all just help ourselves to whatever we wanted, the idea of ‘owning’ things would disappear. Now, by definition, you can’t steal something unless it belongs to someone else. Stealing presupposes that people own things. But people can only own things if they don’t all go around helping themselves whenever they want. So it is logically impossible for everyone to steal things. In other words, it is inconceivable – a contradiction in conception – for everyone to steal things. We can’t conceive of the maxim ‘To steal something I want if I can’t afford it’ being a universal law, so we can’t rationally will it to be a universal law. And so stealing (at least stealing just because one wants something) is wrong. (As argued above, the objection that a world without property involves no contradiction in conception or will is irrelevant.)

However, does this apply to all stealing? For instance, the maxim ‘To steal in order to save a life’ would not, if universalised, lead to the end of property, because it is rare that anyone would need to steal for this purpose.

However, it is unlikely that Kant would agree with this amendment. Our maxim is still ‘To steal’, and it is this that causes the contradiction in conception, so perhaps the purpose for stealing isn’t relevant. Stealing to save a life is still stealing, it is the same *kind of* *action* as stealing because one wants something.

Actually, this last claim can be challenged. Perhaps the maxim of stealing to save a life is actually ‘To save a life, even if by stealing’.

Kant’s Categorical Imperative obviously allows that we should pursue good ends, such as saving lives; indeed, it is a contradiction in will not to help others. But the whole idea of justice and rights is that there are constraints on *how* we pursue good ends. Kant agrees – the duties of justice are more stringent than the duties of virtue (doing good). So we should not steal.

Kant gives a second formulation of the Categorical Imperative, known as the Formula of Humanity: ‘Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end’. This provides another argument that rules out stealing. To steal from someone involves not allowing them to make an informed choice. Why not just ask them to give you what you want? Of course, that is preferable. But what if they refuse? Is it always wrong to steal from someone who, in refusing to help, doing something that is morally wrong?

If we disagree with the answers Kant’s theory gives us, we can use example of stealing to object to the theory. For instance, we can object that in morally bad circumstances, where someone will die or we live in a repressive and unjust regime, stealing is justified, but because Kant’s theory doesn’t recognise this, it gives us the wrong answers.

# Aristotelian virtue ethics

Aristotle’s account of justice is largely deontological. The virtue of justice is defined primarily in terms of doing just actions. Although Aristotle does not understand justice in terms of rights (the concept was invented later by the Romans) or as constraints on what we may do, he comments in the *Nicomachean Ethics* that there are some actions that are never in the mean, but always wrong, and he includes stealing among them. Stealing is always a matter of injustice, of depriving someone of their ‘due’ or ‘fair share’.

What about stealing in an unjust situation, where some people have more than is fair while others are poor? There are two possible responses Aristotle might offer, depending on how the situation came about. Aristotle distinguishes unjust states of affairs from unjust actions. In the first scenario, an unjust state of affairs has come about through no one’s action. This is unfortunate. However, to deliberately choose to do an unjust act, such as stealing, is worse, and so can’t be justified just by appealing to an unjust state of affairs. In the second scenario, the unjust state of affairs has come about through people’s actions, e.g. the rich are deliberately exercising their power to keep the poor poor. This situation demands justice in rectification, to make equal what has been made unequal. Stealing is not normally an act of justice in this sense, but if it were (Robin Hood again?), then it could be justified.

A third response is to disagree with Aristotle that stealing is an act that is never in the mean. We could generalise the argument just made to say that stealing can be justified if it does not involve depriving someone of their ‘due’ or ‘fair share’. When this is so will require practical wisdom, but we should not assume that just because someone owns something that they own it ‘fairly’. Or again, we may argue that even if stealing does (appear to) violate justice, it may be justified by appealing to some other virtue, such as kindness in saving someone’s life. Conflicts between virtues must be resolved by practical wisdom.

# A final thought

A final way of thinking about stealing, which is available to rule utilitarianism, Kantian deontology and Aristotelian virtue ethics, is to revisit property rights. Property rights are not ‘absolute’. I cannot do *whatever* I want with my books, e.g. I can’t start a fire inside a cinema with them, I can’t throw them at people, I can’t withhold them from the government if I haven’t paid my tax, and I can’t keep them if they were stolen from someone else. There are limitations to ownership.

How does this help? To steal something from someone, that person must own it, i.e. their property rights over it must be legitimate. If someone steals your phone, you aren’t stealing if you take it back without their permission. If, in the tales of Robin Hood, the rich did not own their wealth because they had literally stolen it from the poor, then we could say that Robin Hood didn’t steal from the rich at all. He simply returned to the poor what was stolen from them. But perhaps a better description is that the rich did not own their wealth according to *just* laws of ownership, e.g. the taxes that it came from were *clearly* exploitative and so unjust. Robin Hood did steal from them, but if their ownership was unjust, then perhaps his stealing was no violation of justice, and so not wrong.

There remains the question of which theory can best explain this thought.

**2.2. Eating animals**

# Utilitarianism

Act utilitarianism, in its simplest form, says that an action is right if it maximises happiness, and wrong if it does not. Bentham was aware that his identification of happiness – understood as pleasure and the absence of pain – as the only good has some radical implications. One is that animals are morally important. The question about who or what to consider when looking at the consequences of our actions is not ‘Can they reason? nor, Can they talk? but, Can they suffer?’

Utilitarianism says that happiness is good, not just that the happiness of humans is good. If happiness is good, then it is good no matter what creature feels it. There is nothing in the theory that gives us a reason to privilege human happiness over the happiness of non-human animals. So it seems that the logic of utilitarianism requires us to take as much account of beings that are not human as of human beings.

This line of thought has been more recently developed by Peter Singer in his book *Animal Liberation*. We do not think that it is right to treat women worse than men just because they are women (this is sexism), nor to treat one race worse than another (this is racism). Likewise, it is wrong to treat animals as unequal just because they are not human. This is ‘speciesism’.

We can object that with women and men, and different races, there is no difference in those important capacities – reason, the use of language, the depth of our emotional experience, our self-awareness, our ability to distinguish right and wrong – that make a being a person. But there is a difference between human beings and animals with all of these.

Singer responds that these differences are not relevant when it comes to the important capacity that human beings and animals share, namely sentience, the basic consciousness needed to experience pleasure and pain. For a utilitarian, an act (or rule) is wrong if it produces more suffering than an alternative. Who is suffering is irrelevant. When it comes to suffering, animals should be treated as equal to people.

Does this mean that eating animals is always wrong? Not necessarily. First, there is the question of whether stopping the practices of farming, slaughtering and eating animals would reduce the amount of (animal) suffering in the world more than it would increase (human) suffering. It may seem obvious that it would, but the point must be considered carefully. Second, the utilitarian position only objects to suffering, not to killing. Although it can seem very strange to think about it in this way, if you painlessly kill an animal and bring another animal into being, you haven’t reduced the total amount of happiness in the world. According to Singer, we need only ensure that animals are happy when they are alive, and slaughtered painlessly. This would make eating meat much more expensive, because animals would have to be kept in much better conditions. Eating meat is only wrong when animals are not treated as well as they could be. The way in which we rear animals for food at present means that eating meat is wrong.

# Kantian deontology

Kant argues that we should ‘Act only on that maxim through which you can at the same time will that it should become a universal law’ – his ‘Categorical Imperative’. In other words, we should act only on maxims that it is possible for everyone to will. ‘Everyone’ here refers to ‘everyone with a rational will’. On this, Kant argues that there is a sharp distinction between human beings and animals. Animals do not have a rational will. They have desires, but to have a will is to be able to stand back from one’s desires and reflect on whether or not one’s desires are good and whether or not to act on them. And so it seems that any maxim that concerns how we treat animals can be universalised. In particular, there is no contradiction, either in conception or in the will, in universalising the maxim to eat meat.

We find the same result when we turn to the second formulation of the Categorical Imperative, which states ‘Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end’. Because of our capacity for practical reason, human beings are ends in themselves. We have a rational will and can adopt ends. This is the only thing that is unconditionally good, and for everything else that is good, its goodness depends upon being adopted by a will. Animals are not rational wills and are therefore not ends in themselves, and can therefore be treated as means to our ends. We may therefore eat them.

Despite these results, Kant argued that we may not treat animals in any way we want. He starts from the idea that we have the duty to others (and to ourselves) to be virtuous. We have a duty to protect and develop our ability to have a good will and to do our duty. He then argues that if we lack kindness towards animals, we may become unkind towards other people – and this would be morally wrong. Therefore, we need to treat animals in such a way that we don’t damage our own abilities to be virtuous. So while we have no duties to animals, we do have duties concerning animals, but these are indirect duties to ourselves and other people.

We can object that Kant’s theory is very counterintuitive, and fails to account for what is wrong about treating animals badly. Instead of saying that harming an animal is wrong because of the harm done to the animal, Kant says it is wrong because of the harm done to ourselves.

Second, we can note that babies also aren’t rational or autonomous (yet) and neither are some people with severe mental disabilities. Can we treat them as means to an end, or do we have moral duties towards them? We may argue that we do, because babies have the potential to become rational wills, and will do so if cared for. But this argument doesn’t work for people with severe mental disabilities – they will never develop a rational will. If we have duties to them, and yet these human beings do not have different psychological capacities from certain animals, then to deny those animals similar moral consideration would be speciesist, it seems. Whatever is the basis of our duty of care towards people with severe mental disabilities, we may argue, should equally give us a duty of care towards animals.

# Aristotelian virtue ethics

According to Aristotle, we aim at eudaimonia, the good for a human life. It is often translated as ‘happiness’ but Aristotle says it is ‘living well and faring well’. Virtues are traits that help us achieve this. However, animals have no share in eudaimonia, Aristotle argues, because they are incapable of either practical or theoretical reason. Our primary concern with eudaimonia has little place, therefore, for the consideration of animals.

However, more recent virtue theorists have argued that this is mistaken. A different understanding of the relationship between human beings and animals provides arguments for greater concern. There may well be ways of treating animals that are not virtuous. Although she doesn’t phrase it in terms of virtue, one philosopher whose approach has much in common with virtue ethics is Cora Diamond.

## Cora Diamond, ‘Eating meat and eating people’

Diamond argues that eating meat is often wrong, but that arguments about speciesism fundamentally misunderstand ethics. If we start from the idea that we must consider the suffering or happiness of animals and human beings equally, we misunderstand what is important in our relationships with both animals and with other human beings. We can see this by looking at some examples.

First, we don’t eat our dead. This isn’t because we think it will cause them suffering (they’re dead!) or because killing them for food would violate their rights (although it would). It is because a person is not something to eat. That is not true of animals – many animals are ‘things to eat’. It is notable that Singer doesn’t dispute this, but doesn’t notice its significance. He thinks there is nothing wrong with a vegetarian eating a cow that is accidentally killed by lightning. Second, on the thinking of speciesism (equal concern for happiness of human beings and animals), there is no moral difference between having sex with a person of a different race and having sex with a (consenting) gorilla. Yet that is obviously not how we think about people and animals. An animal is not something to have sex with. The capacities of human beings and animals don’t explain why we treat human beings and animals differently. Speciesist arguments are unconvincing because they don’t recognise why we do treat people and animals differently.

We don’t raise people for meat. We don’t number children, but give them names. We celebrate babies being born and people getting married and we give people funerals. We don’t do any of this for animals. This isn’t because all these practices are in the interests of human beings, but because these practices form part of what it is to recognise a human being as the kind of being to be treated with moral concern. Our relations to other people in marking birth, sex and death as serious and important inform the concept of ‘human being’. And so does the idea that we owe them moral duties. Moral duties aren’t the result of the interests of a particular species of animal, homo sapien. Moral duties partly constitute recognising that animal as a human being. It is this recognition, not the thought that we shouldn’t make some being suffer, that is the source of moral life as such. If we are to show that eating animals is wrong, we cannot do so by trying to eradicate the difference between human beings and animals that define those very concepts.

So what reason is there to be vegetarian? We must first recognise that there are many different practices in eating animals, and some may be wrong while others are not. For example, rearing your own pigs, looking after them well and killing them humanely for yourself is very different from picking up sliced pork in the supermarket. To state the obvious, the meat industry is an industry, with animals reared, slaughtered and shipped around the world on an industrial scale, and many of the things that are done to animals in the process are done not because they benefit the animals, but to make the process more efficient.

There are many ways we feel towards animals, and teach children to feel towards animals, that are in conflict with how we treat them in the meat industry. For example, animals feature heavily in nursery rhymes and other stories; we feed birds and squirrels; we stop children from harming pets and respond with horror if they are cruel to them. These emotional responses and behaviours towards animals reflect the idea that we and they are ‘fellow creatures’, not in the sense that we are biologically similar, but in the sense that we are all ‘in the same boat’ as living creatures. Like human beings, animals can die, they lead their lives without our knowledge of what they do, they can provide company for us. These thoughts recognise and respect the independent lives that animals have.

While it is a normal part of thinking of animals as fellow creatures that we eat them, it nevertheless also enjoins us to rear them well or hunt them fairly. But it conflicts strongly with thinking of them as simply part of the production of meat. That way of thinking strips them of respect. We might compare it to thinking of human beings as slaves. In both cases, our capacity to treat the other as independent and to respond with pity and fellow-feeling to how their lives are going is diminished. Animals appeal to our sense of morality not by an assertion of equal interest but by appeal to our pity, to not be callous in how we treat our fellow creatures. This appeal does not try to obliterate the difference between people and animals, but recognises how people may respond to animals.

We can see the basis of Diamond’s approach in virtue ethics in these last points about pity. Just as ethics is grounded in the ‘good life’ for human beings, we recognise – through our emotional responses to other animals – that they share with us the capacity for lives that go better or worse. Not to be emotionally responsive to this is to display the vice of callousness; to treat animals purely as a means to our own ends displays selfishness. As Diamond’s examples of rearing one’s own pigs or hunting them fairly show, killing animals for food may be morally permissible, if it is done for the right reason and in the right way, i.e. with the appropriate feelings. Because that is so often not the case, eating animals is often morally wrong.

**2.3. Telling Lies**

# Utilitarianism

Act utilitarianism, in its simplest form, says that an action is right if it maximises happiness, and wrong if it does not. So a simple act-utilitarian approach to lying would consider whether telling a lie creates greater happiness than telling the truth (or keeping silent). If it does, then it is morally right. If it doesn’t, then it is morally wrong.

John Stuart Mill’s brief discussion of lying in *Utilitarianism* demonstrates that his version of utilitarianism does not evaluate actions just in terms of immediate or obvious consequences, but places them within a bigger picture. A person’s being truthful is of great benefit to people’s happiness generally, and our being able to trust what others say is not only the basis of social well-being but also a foundation of civilisation and virtue more generally. Weakening either our tendency to be truthful or other people’s trust is, therefore, severely damaging to happiness. To tell a lie just for the sake of convenience is therefore morally wrong.

That said, Mill allows that lying is sometimes permissible; e.g. when it is the only way we can withhold information from someone who intends to do harm. We need to carefully consider which situations permit lying by weighing up the conflicting utilities involved. Mill discusses an example from Kant’s ‘On a Supposed Right to Lie from Altruistic Motives’. Suppose someone comes to your house to seek refuge from someone who wants to murder them. Soon after they have hidden, the would-be murderer arrives and asks you where they are. In this case, says Mill, the harm done by lying is outweighed by the good done by saving a life, and lying is permissible.

Rule utilitarianism says that an action is right if, and only if, it complies with those rules which, if everybody followed them, would lead to the greatest happiness (compared to any other set of rules). Rule utilitarians may argue that the rule ‘don’t lie’ will, if everyone followed it, create more happiness than a rule that permitted lying. However, we can object, with Mill, that never lying will lead to harm in certain situations. We need a rule that allows for exceptions. It may be very difficult to put such a rule into words, since the situations in which telling the truth will lead to more harm than good are quite varied. We might lie to prevent someone from doing harm to others; or from doing harm to themselves; or because the truth would hurt (e.g. in cases of terminal illness or sexual infidelity); or because the truth would be damaging to some long-term good (e.g. in politics); or … It is hard to know what the right ‘rule’ for lying should be.

# Kantian deontology

Kant argues that we should ‘Act only on that maxim through which you can at the same time will that it should become a universal law’ – his ‘Categorical Imperative’. If we lie, we are usually following the maxim ‘to tell a lie to get what I want’. If everyone told lies when they wanted to, people would stop believing each other. But you can deceive someone with a lie only if they believe you. So the maxim cannot be universalised, and lying is wrong. Even in the example of the would-be murderer who asks you where his victim is, Kant says you should not lie.

Most people would disagree, and argue that lying in such a situation is the right action. But perhaps Kant is wrong about your maxim in this case. Your maxim may be more specific than lying whenever one wants. For instance, you may adopt the maxim ‘to lie when it is necessary to save a life’. Arguably, this can be universalised. Because in most situations, no lives are at stake, if everyone acted on this maxim, people would still believe each other most of the time. This would mean that in Kant’s example of the would-be murderer who asks you where his victim is, it would be permissible to lie.

There is a question why Kant didn’t simply take this approach when discussing. Why does he argue that even in this case, lying to the would-be murderer is morally wrong? Perhaps Kant might reply if everyone lies when it is necessary to save a life, then the would-be murderer knows this. So they won’t believe us when we answer their question about where their victim is hiding. So we can’t deceive them. The maxim ‘to lie when it is necessary to save a life’ isn’t universalisable after all.

In fact, in his essay on lying, Kant adds a further argument against lying. We don’t know what consequences will follow from our lying. Suppose we lie about the person hiding in our house, saying they ran down the street. Suppose that, unknown to us, they did exactly that. They left their hiding place and ran off. And so our lie sends the murderer straight to where the person is. We would be responsible, Kant claims, for this consequence. If we are tempted to lie because we think the consequences will be better than if we told the truth, it is possible that we are mistaken. We will have failed to do our duty, achieved nothing, and be responsible for the results. It is better to do our duty.

But, we can object, why aren’t we similarly responsible if we tell the truth: if we say where the person is hiding, and the murderer finds them there?

Kant gives a second formulation of the Categorical Imperative, known as the Formula of Humanity: ‘Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end’. This provides another argument that rules out telling lies. To lie to someone is to treat them as a means to our own ends. They are not able to make an informed choice about what to do, but are manipulated in a way that they are unaware of. They can’t share in our ends, because we have not been honest about what our ends are. We should not lie even when the other person’s ends are immoral, and we are trying to prevent those ends being realised. We should not deceive the other person about our intention to thwart their ends. We must give them the chance to share our end of persuading them not to act on their immoral ends.

What seems to follow from Kant’s deontology is that if everyone were morally good, then lying would always be wrong. But sometimes we need to protect ourselves (and others) from the wrong actions of others, and lying may be the only means of doing so. If the action someone intends to do would treat me as a means to an end, then we can, by lying, prevent this result. Kant himself recognises this in his Lectures on Ethics: ‘if I cannot save myself by maintaining silence, then my lie is a weapon of defence’.

# Aristotelian virtue ethics

When Aristotle discusses truthfulness in the *Nicomachean Ethics*, he opposes being truthful to boasting and mock-modesty. So his primary focus is on being truthful about oneself. But he also comments that ‘falsehood is in itself mean and culpable, and truth noble and worthy of praise’. One way of understanding this is to say that lying is an act, like adultery and murder, that has no mean. Lying is already an excess or deficiency in some way, and cannot be virtuous. An alternative interpretation is to say that truth is a final end, something that we should seek not for some further purpose, but for its own sake. This doesn’t entail that lying is always wrong. Pleasure is a final end, but we should not always pursue it – there are appropriate and inappropriate ways of doing so. Perhaps the same can be said of truthfulness.

Aristotle is not particularly critical of boastfulness – to lie about what you have or can do, just because you enjoy lying, is contemptible but ‘futile rather than bad’. To lie in order to gain or protect one’s reputation is not particularly blameworthy, since having a good reputation, in Aristotle’s eyes, is good. Someone who lies to gain money, on the other hand, ‘is an uglier character’. These remarks indicate that there are better and worse motives for lying.

But they also suggest that lying is never virtuous. We might object, however, that as discussed above, there are occasions and motives that justify lying. If there are few strict rules in ethics, as Aristotle says, it is unlikely that lying is always wrong. Instead, we will need practical wisdom to judge when it is justified and when it isn’t. If we seek to deceive someone, to do so virtuously, we would need to do so at the right time, with the right motive, about the right truths, and in the right way.

This last point returns us to the point that there are ways of not sharing the truth other than lying. Perhaps the virtuous person will exhaust all the alternatives first before resorting to a lie.

**2.4. Simulated killing**

Simulated killing is the dramatisation of killing within a fictional context, e.g. in video games, films and plays. It is not merely the description of a killing, as in a novel, but a fictional enactment of killing that the audience or gamer can see and hear. There is a difference – possibly a morally significant difference – between witnessing such a killing and playing the role of the killer. So we will first discuss simulated killing in the context of playing the killer in video games and in acting, and then discuss simulated killing in the context of watching films and plays.

We might wonder whether simulated killing should even be a moral concern. No one is actually killed; no act has been done that violates one’s moral duty. For example, in a video game, all that actually happens is that pixels change. It’s ‘just’ a game.

There are two responses to this line of thought. Obviously, if simulated killing is wrong, it is not wrong for exactly the same reasons that killing is usually wrong. But, first, we need not be concerned just with what is actually done (the simulation). Morality may take a concern with what is being represented (the killing). Is it morally acceptable to create or participate in any representation? While it has become widely socially acceptable to play violent video games, video games involving rape and paedophilia are banned in the UK. And yet we can say, just as truly, that such games are ‘just’ games, and no one is actually raped or molested. Our discomfort with saying this shows that simulations are not necessarily morally neutral just because they are simulations. Second, we can be concerned about the effects of simulated killing both on the people involved and on how they then treat other people in real life.

# 2.4.1. Playing the killer

## Utilitarianism

Act utilitarianism, in its simplest form, says that an action is right if it maximises happiness, and wrong if it does not. In playing a video game, no one is actually harmed in simulated killings, so as long as the gamer is enjoying themselves, there is a gain of happiness. However, could engaging in simulated killing increase the risk of harmful behaviour in the real world? Could it lead to an increased risk of

1. killing
2. aggressive behaviour more generally
3. other forms of antisocial behaviour, e.g. gamers being less responsive to others’ distress, or
4. changes in gamers’ attitudes towards violence in general?

(This last effect, unless such changed attitudes are themselves accompanied by decreased happiness, won’t figure in a utilitarian calculus. However, it is something that virtue theorists will be concerned with – see below.)

Some people think, intuitively, that playing violent games must involve an increased risk of this kind. But the claim is an empirical one, and our expectations are sometimes contradicted by psychological research. In *Ethics in the Virtual World*, Garry Young argues that the evidence is not clear. Some studies on the short-term effects of simulated killing (effects for up to 75 minutes after playing) have indicated that there is an increased risk of aggressive thoughts, emotions and behaviour. However, others found that this increased risk only occurred in people with more violent personalities, while others found that it only occurred in boys, not girls.

There have been very few studies looking at the long-term effects of simulated killing. Some reviews of the evidence have concluded that there is an increased risk of aggressive thoughts, emotions and behaviour and a decrease in empathy, but a number of the studies have been challenged as invalid or found an effect so weak as to be insignificant. There is some evidence that journals are also more likely to publish studies that find a link than studies that don’t, so there is a bias in the published evidence. Therefore, the evidence that simulated killing leads to more aggressive behaviour, etc., is unclear, though perhaps we can say that there is an increased risk for some people.

Act utilitarians don’t just consider the actual consequences of an action. They consider the ‘tendency’ or probability of the action having certain consequences. Rule utilitarianism says that an action is right if, and only if, it complies with those rules which, if everybody followed them, would lead to the greatest happiness (compared to any other set of rules). Rule utilitarians consider the consequences of a rule that allows simulated killings. The evidence so far is that we cannot say that simulated killing will probably increase actual immoral behaviour.

However, even if simulated killing increased aggressive behaviour, utilitarians will weigh the decrease in happiness that results from such behaviour in the real world against the pleasure derived from playing the game. Simulated killing will only be wrong if, taking both the happiness and unhappiness caused into account, it leads to less happiness on balance than not engaging in simulated killing.

Are we mistaken in trying to apply the utilitarian calculus to the act of simulated killing directly? John Stuart Mill argues that, in most cases, we only need to consider the ‘secondary principles’ of common morality. But common morality doesn’t provide an obvious guide here, given that video games of this sort have not been around very long. If we look to other games, such as children’s play (cops and robbers, aliens, monsters), simulated killing is widely permitted and considered part of normal development (at least for boys).

Some people, therefore, might condemn playing violent video games as ‘childish’ behaviour that adults would be expected to outgrow. But the utilitarian force of such an objection is unclear. Does engaging in childish play decrease happiness? Perhaps an appeal to Mill’s distinction between higher and lower pleasures adds some weight. Childish pleasures, such as those involved in simulated killing, will not count as higher pleasures for adults. Hence we may think worse of such a person who engages in such activity, but we will not condemn the activity itself.

## Kantian deontology

Kant argues that we should ‘Act only on that maxim through which you can at the same time will that it should become a universal law’ – his ‘Categorical Imperative’. He also expresses the principle as ‘Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end’. Because of our capacity for practical reason, human beings are ends in themselves. We have a rational will and can adopt ends. These principles define our duties. Playing a game per se is no violation of one’s moral duty. But if doing so damages one’s rational will or leads to neglecting or violating one’s duty to other people, then we can object. We could argue that cultivating cruelty and an indifference to virtual suffering through simulated killing could undermine our willingness and ability to treat others as ends in themselves in real life. Kant notes that, at the time he was writing, butchers and doctors were not allowed to serve on English juries because they were hardened to suffering and death.

However, having reviewed the empirical evidence, it seems that there is not enough evidence to say that there is a link between simulated killing and neglecting one’s duties to others.

Even if we don’t fail in our duties to others, perhaps we somehow fail in our duty to ourselves. We could argue that repeatedly engaging in simulated killing erodes our sense of identity as rational, moral beings. But again, it is unclear whether this is true. If it does, then this would be a reason for thinking that it is wrong.

## Virtue ethics

A similar concern is central to Aristotelian virtue ethics. According to Aristotle, we aim at eudaimonia, the good for a human life. It is often translated as ‘happiness’ but Aristotle says it is ‘living well and faring well’. Virtues are traits that help us achieve this. We develop virtues by doing virtuous acts. For instance, we become just by doing just acts. Likewise, we become unjust by doing unjust acts. Killing is often an unjust act. So the cumulative effect of playing games which involve simulated killing may lead to the development of character traits that are not virtuous, such as injustice and unkindness, or at least inhibit the development of character traits that are virtuous, such as justice and kindness. Simulated killing is wrong if it prevents the development of virtue, and so prevents the gamer from achieving eudaimonia.

Aristotle may be right that doing unjust acts develops the vice of injustice. But simulated killing is not an unjust act – no one is killed. So why think that simulating unjust acts will develop injustice? Once again, we can argue that the evidence doesn’t support this claim.

Rather than focus on the development of character, we can ask whether a virtuous person would engage in playing video games that involve simulated killing. If so, then they will do so in the right way, with the right motive, and at the right times, as this is what defines virtuous action. What might that involve? For example, why would someone want to simulate killing someone else? Is taking pleasure in this activity virtuous?

Clearly, there is pleasure to be gained from violent video games – otherwise, they would not be so popular. But there may be more than one kind of pleasure one can take, and more than one motive for killing someone within the game. The virtuous person enjoys such pleasures appropriately, if there is an appropriate way to enjoy them. There may be morally better and worse ways of relating to simulated killing within the game. Is the point of the game just to kill people, or within the narrative, is killing a necessary means to some further goal? Does the gamer enjoy simulated killing as part of doing well in the game (so the motive is competitiveness) or just enjoy simulated killing for its own sake? And so on.

There may also be morally better and worse ways of understanding the relationship between the game and reality. It can be wrong for someone who confuses the two to play the game, but okay for someone who doesn’t. Virtue ethics recognises that the right thing to do is not the same for everyone. The ‘mean’ is relative to each person and relates to each person’s stage of moral development. If someone is at all likely to think of the game world as a model for the real world, playing such games is not virtuous for that person. Someone else could experience the rush of adrenalin as a helpful and safe expression of natural human aggression (good), while someone else again indulge in fantasies of actual killing during play (bad). And so on.

Relating this back to the empirical evidence: someone who draws a clear conceptual and emotional distinction between simulated killing and real life may be at no risk of being more aggressive after playing or developing bad character traits. Someone who cannot draw such a distinction may be at risk, and so should not play.

However, these remarks don’t settle the question of whether the virtuous person would want to play such a game.

## 2.4.2. ACTING THE KILLER

We can develop the points just made in relation to the actors in a film or play. Acting takes place within a context which is governed by a whole set of conventions about what particular actions mean. Arguably, actors don’t imitate real-life killings, and even in films, which may be more lifelike with special effects, etc., violence is typically unrealistic. Instead, actors pretend to kill (and to die) on the understanding that certain actions are to be understood as killings. Furthermore, actors – even method actors – are not supposed to feel genuine lethal rage towards their fellow actors during the scene, nor genuine bloodlust and excitement. (Method actors may feel the fictional counterparts of such emotions ‘in character’, but would not feel such emotions as themselves.) Suppose an actor confessed to feeling real murderous rage after the play or filming. This would be disturbing, to both them and us. Such feelings are not part of the conventions of acting, and indicate a blurring, in the actor’s psychology, between the character and the actor.

What these remarks are meant to show is that acting takes place in a complex social context that sets acting apart from reality. The conventions protect the actors, enabling them to do their job without damaging themselves. Concerns about such a blurring are at issue in the discussion of playing video games: does the gamer fail to distinguish themselves from their character? Is their moral goodness compromised by the immorality of their avatar, either during or after the game?

# 2.4.3. An audience’s perspective

Is there anything morally wrong with watching violent TV shows, films or plays? Such works are fictions, and it is common to talk about ‘suspending disbelief’ when immersed in a film or play. We ‘make-believe’ that what we are seeing is real. We don’t believe it is – that would lead to very different emotions and actions (call the police!). But we pretend or imagine that it is real. Is it wrong to do this when what one is witnessing is a simulated killing?

Although we haven’t discussed this, it is worth noting that on each of our theories, killing is sometimes morally right, e.g. in war (Aristotle) or euthanasia (utilitarianism) or capital punishment (Kant). If a dramatic work explores this issue carefully, and convincingly presents a killing as the morally right thing to do, then it is hard to see what is wrong with imagining the simulated killing (at least on the assumption that such a killing would be morally right). So for the purposes of argument, let’s assume that the killing that is simulated would be morally wrong if done in real life.

The approaches of our three theories to this question have been laid out above, so we can be brief. A utilitarian will be interested in the effects on the overall happiness of watching make-believe killings (or of a rule that allows simulated killings in TV shows, films and plays). There is no immediate decrease in happiness if the audience gets something positive out of the experience. So concerns will be limited to the longer term effects. As we might expect, the evidence is very similar to the evidence connecting playing violent video games to aggression in real life. There is some evidence that the link is stronger in some groups of people than others, but overall, we don’t have enough evidence to conclude that, in general, watching violence on screen or on the stage is likely to make one a less moral person. Even if there were a link, the risk of diminishing happiness needs to be weighed against the enjoyment gained by watching such works.

Let us take deontological and virtue ethical concerns together. Irrespective of consequences for how one acts, does watching simulated killing damage one’s character or good will intrinsically? Immoral simulated killings can take place within two fictional contexts. In one, the killing is represented as immoral, the killer as morally or emotionally wretched. The work can be understood as a morality tale – this is how not to be. But the killing can also be represented as moral – the morality of the work is different. This is the most problematic case. Is it wrong to imagine that something that is immoral is actually moral?

Kant would argue that it is certainly irrational. What is immoral cannot be moral. We can coherently imagine that contingent truths are different. But moral truths are established by a test of what is possible – so they are not contingent, but necessary. But is there anything morally wrong with imagining something impossible?

Again, understanding the relation between the work and moral reality is important. For example, do we think that the author intends the (immoral) values to be moral values only within the fiction? Or is the message that we should live according to the values portrayed? On the one hand, there may be something not virtuous about joining with the immoral imagination of the author. On the other hand, one may argue that it can help one understand morality more deeply. But this will only occur if one can keep one’s distance from the ‘morality’ of the dramatisation. A virtuous person will be alive to the moral implications of the story being told, not simply in terms of its effects but in terms of its representation of what a good life is and the place of killing within it. Their make-believe will be coloured by this awareness.

3. Meta-ethics

We usually think of ethics as about how we should act and live. This is the approach taken by normative ethical theories, which provide an account of which actions, motives and character traits are right or good. They are intended to provide guidance on how to live. Metaethics, by contrast, does not do this. It asks about what morality is, philosophically speaking. It asks questions in philosophy of language, philosophy of mind, metaphysics, and epistemology.

1. Philosophy of language: what do statements like ‘Murder is wrong’ or ‘Courage is good’ mean? Are these statements of fact? Can ethical claims be true or false? Or are they something else, such as expressions of our approval or disapproval of certain actions or character traits?
2. Philosophy of mind: what is it to hold a particular moral view, e.g. that murder is wrong? If ‘murder is wrong’ states a truth, then moral views are factual beliefs. On the other hand, if ‘murder is wrong’ expresses a feeling, then moral views are attitudes of approval or disapproval (or something similar). Which theory is correct? One relevant issue is whether holding a moral view is a matter of being motivated to act in certain ways, e.g. not to murder. If it is, what does this imply about the nature of morality?
3. Metaphysics: suppose we think that ethical language states truths. Are these truths objective? Are there moral properties, like being right or wrong, that are part of reality?
4. Epistemology: if there are ethical truths, how do we discover what these truths are? On the other hand, suppose we deny that ethical statements are true or false, arguing that they are expressions of subjective feeling. In that case, is there such a thing as moral reasoning? Can we provide reasons that justify our actions?

# The origins of moral principles: reason, emotion/attitudes, or society

One way to begin thinking about these questions is to ask where our views about what is right and wrong, good and bad, come from. Now, of course, in one straightforward sense, our moral principles come from the people we knew when we were children, our parents, teachers and so on. The same is true of a great many of our beliefs and preferences. The origin of my belief that the sun is 93 million miles away from the Earth is my physics teacher at school. The origin of my taste in music is the group of friends I had as a teenager. The origin of my moral principles is my parents. And so on. But this kind of answer only says what the cause of my belief is. It is particular, not general, because it only explains my belief. And it is contingent, because something else could have caused my belief. Perhaps my taste in music could have come from my physics teachers and my belief about the distance of the sun from my teenage friends!

Suppose we could generalise from just my beliefs and tastes to where people’s beliefs and tastes in general come from. Still, in asking about the origin of our moral principles, we aren’t looking for a purely causal, particular or contingent answer – an answer that psychology or sociology might provide. We are interested in the ultimate origin of anyone’s moral principles, and we want a philosophical answer that tells us something about the nature of morality. Just talking about how someone, or even most people, acquires their moral principles doesn’t help for two reasons.

First, if I got my belief about the distance of the sun from my physics teacher, where did he get his belief from? If he got it, in turn, from his teacher, we can ask the question again – where did his teacher get his belief from? And so on. This is a regress and uninteresting. Much more interesting, philosophically, is what’s the ultimate origin of the belief? How do people arrive at beliefs of this kind in the first place? In this case, it is a set of experiments and calculations conducted by scientists (in ancient Greece, Aristarchus; in modern times, Christiaan Huygens and Giovanni Cassini). We could, if we wanted, repeat these experiments and calculations for ourselves. This tells us that the origin of the belief is in scientific investigation – a much more informative answer than ‘my teacher’, and one that offers a justification of the belief, not only a causal story.

Second, our beliefs about the distance of the sun is objectively true or false, but our tastes in music aren’t about truth at all – yet (let’s suppose) we acquire both from people we knew in childhood. And while our beliefs about the distance of the sun originate in scientific investigation, the same is not true of our tastes in music! Saying that we also acquired our moral principles from people we knew in childhood is uninformative, because it doesn’t distinguish between different kinds of psychological state. For instance, it doesn’t say whether moral principles are more like beliefs about the distance of the sun or musical tastes or something else again.

Different forms of ‘moral realism’ argue that there are moral truths, and we can discover these truths by using reason. On some versions, the reasoning is largely empirical, so the origin of moral principles is somewhat similar to the origin of scientific beliefs in rational investigation of the natural world. On other versions of moral realism, we must use rational intuition, so the origin of moral principles is somewhat similar to the origin of mathematical beliefs in a priori reasoning.

The claim that our principles have their origins in emotions or attitudes is defended by versions of ‘non-cognitivism’ (see below), including ‘emotivism’ and ‘prescriptivism’. On these views, we don’t discover moral truths using reason. Our moral principles aren’t about truth at all; they are expressions of how we feel and how we want ourselves and other people to act. Our moral principles originate in our emotions and attitudes (even if these emotions and attitudes are influenced by other people’s emotions and attitudes).

‘Moral relativism’ claims that the origin of moral principles is society. We will discuss this theory only in passing, but it is important to be aware of it. Moral relativism argues that morality originates in how a society regulates the relationships between people. Moral principles are not expressions of how individuals feel, because they are essentially social, shared. But neither are they discovered by reason, because there are no truths about which moral principles societies should have. There are just the moral principles that societies in fact have. Within any society, moral principles record how that society says people should behave. They are essentially social, arising and evolving in a social context through interaction with how the society itself changes over time.

# The distinction between cognitivism and non-cognitivism

Theories in metaethics fall into two broad families – cognitivism and non-cognitivism. The distinction is now understood by philosophers to depend on whether one thinks that moral judgements express beliefs or not.

Cognitivism claims that ethical language expresses beliefs. Beliefs can be true or false, so ethical claims that can be true or false. To believe that murder is wrong is to believe that the sentence ‘Murder is wrong’ is true. Because (usually) a claim is true because it correctly describes how the world is, cognitivists (usually) also claim that ethical language aims to describe the world.

Non-cognitivism claims that ethical language does not express beliefs, but some other, non-cognitive mental state. And so ethical claims do not try to describe the world and cannot be true or false. Different non-cognitivist theories disagree on exactly what kind of mental state is expressed by moral judgments, but it is usually an attitude or feeling. So ‘Murder is wrong’ is neither true nor false, but an expression of, say, the speaker’s disapproval of murder.

## Mental states and ‘direction of fit’

In her book *Intention*, Elizabeth Anscombe explained the difference between a cognitive mental state and a non-cognitive mental state in terms of the idea of ‘direction of fit’. A man goes shopping, taking his shopping list with him. When shopping, he uses his list to guide what he puts in his basket. At the end of the shop, what is in his basket should ‘fit’ his list. If it doesn’t, the mistake is with the basket, and the basket should be changed to fit the list. Now suppose that the man is being followed by a store detective. She makes a list of each thing that the man puts in his basket. At the end of the shop, her list should ‘fit’ his basket. If it doesn’t, the mistake is with her list, and the list should be changed to fit the basket.

The shopper’s list is a list of what he wants. Desires have a ‘world-to-mind’ direction of fit. We seek to change the world to fit our desires and thereby satisfy them. They are not true or false, but represent how the world should be. By contrast, the detective’s list is a list of what she believes is in the shopper’s basket. Beliefs have a ‘mind-to-world’ direction of fit. We change our beliefs to fit the world, and thereby have true beliefs. They represent how the world is, not how we want it to be.

So which direction of fit do moral views have? Is the thought ‘murder is wrong’ a belief about how the world is, or is it like a desire to make the world a place in which there is no murder? Both answers are plausible and both answers face challenges.

## Issues

Non-cognitivists argue that moral judgements are, like desires, motivating. Holding the view that murder is wrong involves being motivated not to murder. But, they continue, factual beliefs are not motivating. The sun is 93 million miles from the Earth – so what? Believing that fact inclines me to do nothing in particular at all. Because moral views are motivating, they are not beliefs, but non-cognitive attitudes.

Cognitivists can respond that some beliefs, including moral beliefs, are motivating. Or they can argue that moral beliefs aren’t motivating. Instead, caring about what is morally good or right is motivating. It is possible, therefore (but perhaps psychologically very unusual), to believe that murder is wrong and not be motivated to refrain from murdering because one simply doesn’t care about morality.

Cognitivism argues that what is right or wrong is something we can be mistaken about. It isn’t just ‘up to us’ whether murder is wrong. People who think that murder is just fine are mistaken and vicious. Morality isn’t simply a matter of taste. Non-cognitivism, therefore, faces the challenge of explaining why we make a distinction between morality and personal taste. Is non-cognitivism going to lead to nihilism about morality, the view that nothing is right and wrong?

Non-cognitivism can argue that it is a simpler theory. It has a simpler metaphysics and a simpler epistemology. Cognitivism needs to explain how moral claims can be objectively true or false. Are there moral properties ‘in the world’? What kind of property could they be, and how can we find out about them?

# Moral realism

## From cognitivism to moral realism

Cognitivism is the view that ethical language expresses ethical beliefs about how the world is. Cognitivists argue that moral judgements can be true or false, and so aim to describe the world. Furthermore, we can be mistaken about whether a moral judgement is true or false. Our thinking it is true does not make it true.

Here are three quick arguments in favour of cognitivism:

1. We think we can make mistakes about morality. Children frequently do, and have to be taught what is right and wrong. If there were no facts about moral right and wrong, it wouldn’t be possible to make mistakes.
2. Morality feels like a demand from ‘outside’ us. We feel answerable to a standard of behaviour which is independent of what we want or feel. Morality isn’t determined by what we think about it.
3. Many people believe in moral progress. But how is moral progress possible, unless some views about morality are better than others? And how is *that* possible unless there are facts about morality?

But if there are truths about morality, what kind of truths are they? Moral realism claims that good and bad are properties of situations and people, right and wrong are properties of actions. Just as people can be 5 feet tall or run fast, they can be morally good or bad. Just as actions can be done in 10 minutes or done from greed, they can be right or wrong. These moral properties are a genuine part of the world. Whether moral judgements are true or false depends on how the world is, on what properties an action, person or situation actually has.

Moral realism in the last 150 years has focused on trying to clarify the precise nature of the relation between moral properties and natural properties. This has led to two positions: moral naturalism and moral non-naturalism. Moral naturalism claims that moral properties are natural properties; moral non-naturalism claims that they are a distinct, non-natural kind of property.

The debate is important because it has significant implications for our understanding of both philosophy and morality. Philosophy, first. Moral claims are not analytically true. That a particular action of killing someone, say, is morally wrong is not something that is true by definition of the concepts involved. So if moral claims are true at all, they must be synthetic propositions. Now, if we think that empiricism is correct, then we could only gain knowledge of moral judgments through empirical investigation, i.e. sense experience and scientific investigation. If that is possible, then moral properties must be natural properties. But is it possible? Could we really learn the difference between right and wrong through sense experience? Could science improve or correct our ethical views? Perhaps this sounds rather odd, and we don’t learn about morality in these ways. Then if realism is true, we must gain moral knowledge in some other way, and that would mean that empiricism is false. It would also mean that moral properties are not natural properties, and so then there is more to the world than what can be investigated by science. So the debate between moral naturalism and moral non-naturalism has significant implications for the debate between rationalism and empiricism and for our view of what exists.

If we can show that moral naturalism is false and that moral non-naturalism is false, then we have shown that moral realism is false. That means that we face the question of whether there are *any objective moral truths*. If morality isn’t objective, we may think that has serious implications for how we live our lives. For instance, why bring up children not to steal and not to lie if it is just a subjective matter whether these things are wrong or not? We don’t force children to play particular sports – which sports they enjoy and pursue is up to them. If there is no objective morality, shouldn’t we do the same with stealing and lying? We can draw similar implications for the criminal law and punishment. If there is nothing objectively wrong with murder, should we imprison someone for committing it? Some of the technical debates in metaethics can seem distant from our everyday concerns, but issues such as these lie behind them.

# Moral naturalism

Moral naturalism claims that moral properties are natural properties. But what counts as natural? Because there is disagreement on the answer to the question, there is more than one type of ‘moral naturalism’, which we can call ‘reductive’ and ‘non-reductive’.

Many philosophers accept the definition of ‘natural properties’ that was given by G. E. Moore, namely properties that we can identify through sense experience and science. On this definition, moral naturalism is a form of reductionism. It claims that the things in one domain – moral properties of goodness and rightness – are *identical* with some of the things in another domain – certain properties that we can identify through sense experience and science. The most plausible natural properties that might be moral properties are certain psychological properties, e.g. happiness. The identity claim is a reduction because we have ‘reduced’ moral properties – which we might have thought were a different kind of thing – to psychological properties. I.e. there is *nothing more* to moral properties than being a certain kind of psychological property.

Non-reductive moral naturalism argues that morality is an expression of the natural capacities of human beings, the capacities we have as a species of animal, part of nature. This kind of naturalism wants to reject ‘supernatural’ explanations of morality, e.g. that what is right or wrong is determined by God (as some theories in religious ethics claim) or that when making moral judgments, we use some ‘non-natural’ faculty of reason (as Kant thought). Moral properties are a kind of natural property, but they can’t be reduced to some other kind of property, such as the properties that science investigates.

The difference between reductive and non-reductive moral naturalism may become clearer as we look at examples of each.

## Utilitarianism as naturalism

The normative ethical theory utilitarianism claims that the only good is happiness. This can be interpreted, in metaethics, as a form of reductive claim. We can interpret this to mean not simply that happiness is the only thing that is good, but that happiness is what goodness is. They are the same property. Happiness is a natural (psychological) property, and therefore, so is goodness. Because happiness is a natural property, so is maximising happiness. Whether an act maximises happiness is a (complex) natural property. According to utilitarianism, an act is right if it maximises happiness. Therefore, rightness is also a natural property. On this interpretation, utilitarianism is a form of reductive moral naturalism.

Bentham appears to understand utilitarianism in these ways. For example, he opens *An Introduction to the Principles of Morals and Legislation* by saying that pleasure and pain not only determine what we *ought* to do, they determine what we *shall* do. This is a psychological claim: we are only ever motivated by pleasure and pain. Utilitarianism starts from recognising this natural fact, and builds on it.

This is how Moore in *Principia Ethica* interprets Bentham as well as Mill’s ‘proof’ of utilitarianism. In his ‘proof’, Mill argues that what is good is what is desirable. We can find out what is desirable by finding out what people desire. He then argues that happiness is desired. From this, he infers that happiness is good. This only works, says Moore, if Mill thinks that what is good is the same as what is desirable, and that what is desirable is the same as what is desired. So Mill must be thinking that goodness is the natural property of what is desired.

Moore goes on to argue that, in making the argument, Mill commits the fallacy of equivocation in this argument, confusing two meanings of a word. The word ‘desirable’ has two meanings. Its usual meaning is ‘worthy of being desired’. Anything desirable in this sense is good. But another meaning could be ‘capable of being desired’. To discover what is capable of being desired, look at what people desire. Mill links what is desirable to what people desire. But what people actually desire is not the same as what is worthy of being desired (good). People want all sorts of rubbish! Mill has assumed that what people desire just is what is good; he hasn’t spotted that these are distinct meanings of ‘desirable’.

But Moore’s objection misinterprets Mill’s argument. Mill is asking ‘What evidence is there for thinking that something is worthy of being desired?’ He argues that people in general desire happiness. Unless people in general desire what is not worth desiring, this looks like good evidence that happiness is desirable. Is there anything that everyone wants that is not worth wanting? If we look at what people agree upon in what they desire, we will find evidence of what is worth desiring. Everyone wants happiness, so it is reasonable to infer that happiness is desirable (good).

Mill takes what people desire (which is a natural property) as *evidence* for what is desirable (good). He does not say that goodness is the same property as being desired. And when he claims that what is good is what is desirable, nothing he says implies that he thinks that ‘being desirable’ (as opposed to being desired) is a natural psychological property. So we simply can’t say whether Mill is a reductive naturalist or not.

However, there is some evidence of reductive naturalism in Mill’s claim, at the end of the proof, that ‘to think of an object as desirable … and to think of it as pleasant are one and the same thing’. So for something to be good is for it to be desirable, which is the same as being pleasant, which is for it to contribute to happiness. This sounds very much like what Bentham says as well – to say that something is good and to say that something produces happiness is to say the same thing. If the words ‘good’ and ‘produces happiness’ have the same meaning, goodness and (producing) happiness are the same property (just like ‘bachelor’ and ‘unmarried man’ mean the same and being a bachelor is the same property as being an unmarried man).

Whether or not Bentham and Mill are reductive naturalists, a reductive naturalist interpretation of utilitarianism can argue that it solves some of the issues facing cognitivism. If goodness is just happiness, then there is no puzzle about what kind of thing goodness is. Furthermore, we can discover what creates happiness empirically. So we have an answer to the question about how we find out what is morally right and wrong: through experience. We can explain how morality can be objective in terms of what, objectively, contributes to people’s happiness.

It is true that *if* reductive moral naturalism is true, we can discover moral properties through experience. But this fact won’t help show that reductive moral naturalism *is* true. Given the different normative ethical theories that exist, claiming that goodness *is* happiness (or any other natural property) is obviously contentious. Such a claim isn’t something that we can demonstrate by empirical reasoning – no scientific experiments will show that goodness is, after all, just happiness. So from the psychological facts *alone*, we cannot deduce any moral knowledge. We have to defend the identity claim philosophically.

## Naturalism in virtue ethics

Some philosophers have read Aristotle’s virtue ethics as a naturalist theory. They argue that he analyses eudaimonia in terms of the ‘characteristic activity’ (or ‘function’) of human beings as rational beings. The facts about human nature, in particular psychological facts about our desires, our needs and our ability to reason, are the basis for moral truths. There are facts about what our characteristic activity is, there are facts about what traits enable us to perform our characteristic activity well. And so it turns out to be a psychological fact whether a character trait, such as courage or being short-tempered, is good or bad.

In her article ‘Virtue ethics’, Julia Annas rejects this and argues that Aristotelian virtue ethics is a form of non-reductive moral naturalism. Morality is *based on* natural facts about human nature. The sciences of biology, ethology and psychology can help us to identify the patterns of flourishing for each species, and so help us understand what eudaimonia involves for us. Furthermore, it is a *natural* fact about human beings that we are rational animals. Rationality is a natural capacity of human beings, it characterises us as the species of animal that we are. However, virtue ethics can’t be *reduced* to claims about natural facts. The rationality involved in practical wisdom is not just a matter of knowing and applying facts that we can discover through sense experience and science. As rational animals, we create and evaluate ways of living, rather than simply live according to a set pattern. We look at the *reasons* for living a particular way. The person who has practical wisdom is not simply able to grasp some psychological fact about a situation that someone without practical wisdom cannot grasp. Instead, they understand the *reasons* for feeling, choosing or acting a certain way in a certain situation. This is why virtue is in accordance with ‘right reason’.

For Aristotelian virtue ethics to be a form of reductive moral naturalism, we would have to claim that whether some consideration is a reason or not is itself a natural property.

# Moral non-naturalism

Moral non-naturalism claims that moral properties are not natural properties. There are different kinds of moral non-naturalism. We will look at just one, Moore’s intuitionism. Moore understands ‘natural property’ to mean a property that we can discover through sense experience and the sciences, including psychology.

## The naturalistic fallacy

Reductive moral naturalism claims that moral properties are identical to natural properties (of the kind that can be discovered by sense experience and science). In *Principia Ethica*, G. E. Moore argued that moral properties are not natural properties. Moral properties may be correlated with certain natural properties, but they are not identical. Correlation is not identity. For example, having a heart is correlated with having kidneys – every animal that has a heart has kidneys and vice versa. But hearts and kidneys are not the same thing! Or again, having a size and having a shape are correlated – everything that has a size has a shape and vice versa. But size and shape are distinct properties. So even if goodness is correlated with happiness or pleasure, say, that does not show that they are the same property.

Moore called the attempt to identify goodness with any natural property the naturalistic fallacy. To see this, we need to think more about goodness. Goodness, Moore argued, is a simple and unanalysable property. It cannot be defined in terms of anything else. Of course, we can say how people use the term ‘good’, what they apply it to or again, what has the property of goodness. For instance, it makes perfect sense to say that pleasure is good in this sense. But this is to accept that there are two things here, not one. There is the pleasure, and pleasure has this additional property, goodness. Compare: when we say ‘You weigh 60 kilos’, we attribute you with the property of weighing 60 kilos. We don’t think that you are the same thing as that weight – you are a person, not a weight! Likewise, we can meaningfully say that pleasure is good if we distinguish between pleasure and goodness. But we can’t give a definition that defines goodness in terms of its parts that together ‘make up’ goodness.

Colours are similar. We can say what things are yellow, e.g. the sun, ripe lemons, etc. but these things don’t define the colour yellow. Yellow is a simple property, and no one can explain what yellow is to someone who doesn’t know. You have to see it for yourself to understand what it is. (For instance, we can’t define yellow – which is part of our visual experience of the world – in terms of wavelengths of light. It might be correlated with these, such that seeing yellow is always caused by certain wavelengths of light. But it is a mistake to think that they are one and the same thing. Unlike wavelengths of light, colours are conceptually related to vision.)

Unlike colours, goodness is not a natural property. It cannot be investigated by empirical means. It is real, but it is not part of the natural world, the world of science. So, because goodness cannot be analysed in terms of any other property, it is a mistake to think that the property of goodness is identical with any natural property.

## The ‘open question’ argument

Moore supports his view that a definition of goodness is impossible by the ‘open question’ argument. An open question is a question to which the answer could be more than one thing, for instance, it could be ‘yes’ or ‘no’. If goodness just is pleasure, say, then it wouldn’t make sense to ask ‘Is pleasure good?’ This would be like asking ‘Is pleasure pleasure?’ This second question isn’t an open question, because the answer has to be ‘yes’. It cannot, logically, be ‘no’ Put another way, we can say that asking ‘It is pleasurable, but is it pleasurable?’ is a closed question, rather like ‘He is a bachelor, but is he an unmarried man?’ Now, if goodness is the same thing as pleasure, then ‘It is pleasurable, but is it good?’ is also a closed question. But, says Moore, this isn’t a closed question – the answer can logically be ‘yes’ or ‘no’. The same is true of ‘Is pleasure good?’ And so goodness cannot be pleasure, or any other property. ‘Is x good?’ is always an open question while ‘Is x x?’ is not. And so goodness cannot be defined as any other property.

## Is the ‘naturalistic fallacy’ a real fallacy?

Moore’s open question argument doesn’t work. Here is a similar argument. ‘The property of being water cannot be any other property in the world, such as the property of being H2O. If it was then the question “Is water H2O?” would not make sense – it would be like asking “Is H2O H2O?” So water is a simple, unanalysable property.’ This is not right, as water just is H2O.

The reason the argument doesn’t work is because it confuses concepts and properties. Two different concepts – ‘water’ and ‘H2O’ – can pick out the same property in the world. Before the discovery of hydrogen and oxygen, people knew about water. They had the concept of water, but not the concept of H2O. So they didn’t know that water is H2O. ‘Water is H2O’ is not analytically true. However, water and H2O are one and the same thing – the two concepts refer to just one thing in the world. Water is identical to H2O.

Likewise, the concept ‘goodness’ is a different concept from ‘happiness’. ‘Happiness is good’ is not an analytic truth. We can accept that Moore has demonstrated this. But perhaps the two concepts refer to exactly the same property in the world, so that goodness is happiness. Moore’s open question argument does not show that they are different properties.

# Intuitionism

If moral properties are not natural properties, then how do we discover them? How do we know what is good? Mill claims that we cannot prove what is good or not. To prove a claim is to deduce it from some other claim that we have already established. Moore agrees. But unlike Mill, he does not think that we can argue inductively from evidence either. All we can do is consider the truth of the claim, such as ‘pleasure is good’, itself. Moore calls such claims ‘intuitions’.

What does this mean? The claim that some truths can be known by rational ‘intuition’ is made by rationalism. But what is a moral intuition, and how can we tell if it is true? Moore leaves these questions open: *‘when I call such propositions Intuitions, I mean merely to assert that they are incapable of proof; I imply nothing whatever as to the manner or origin of our cognition of them’*. However, he has already said more than this. He has argued that these claims are not analytically true. And he has argued that we cannot know them through empirical investigation. So they must be some variety of synthetic a priori knowledge. He claims that we can know propositions about what is good to be true (or false) by considering the proposition itself. Intuitions are ‘self-evident’ propositions.

A self-evident judgement rests on the ‘evidence’ of its own plausibility, which is grasped directly. This doesn’t necessarily mean that everyone can immediately see that it is true. ‘Self-evident’ is not the same as ‘obvious’. Our ability to make a self-evident judgement needs to develop first, and we need to consider the issue very carefully and clearly. Because moral intuitions are not known through the senses, the self-evidence of a moral intuition will be more like the self-evidence of a necessary truth, such as mathematics or claims about what is logically possible, than the self-evidence of a perceptual truth, such as the claim that there is a table in front of me.

So, intuitionism does not need to claim that we have a faculty of intuition that ‘detects’ whether something is good or not, a bit like a supernatural sense. Intuitionism is simply a form of ethical non-naturalism that claims that some of our moral judgements are synthetic yet self-evident.

**OBJECTIONS TO MORAL REALISM**

In this section, we discuss two objections from David Hume’s *Treatise of Human Nature* to any form of moral realism – indeed, to any form of cognitivism.

# Hume’s argument from motivation

The first argument Hume presents that we shall discuss is this:

P1. Moral judgements can motivate actions.

P2. Reason cannot motivate action.

C1. Therefore, moral judgements are not judgements of reason.

Cognitivism claims that moral judgements express beliefs, which can be true or false. And the faculty of judging what is true or false is reason. Hence, Hume’s conclusion is a rejection of cognitivism.

Hume assumes (P1) to be true. His argument for (P2) depends on Hume’s ‘fork’, the claim that we can have knowledge of just two sorts of claim: relations of ideas or matters of fact. Relations of ideas can be discovered just by thinking, by recognising the truth of an analytic proposition or by deductive reasoning. To deny a relation of ideas is to contradict oneself. Matters of fact are claims about what exists, and they are established by sense experience and causal inference. Hume understands the faculty of knowledge as ‘reason’. So because there are just two types of knowledge, reason makes just two types of judgment – judgments about relations of ideas and judgments about matters of fact. Neither relations of ideas nor matters of fact are motivating.

Hume argues earlier in the *Treatise* that we are always motivated by our emotions and desires. But, he claims, emotions and desires are not psychological states that can be true or false. They don’t show us how the world is, they motivate us to act on it, to change it. By contrast, judgements of relations of ideas and matters of fact show us how the world is. A psychological state that simply presents a truth can’t motivate us to act, because there is no pressure to change the world to fit the mind. Simply understanding that some relation holds between two ideas doesn’t entail that we should act one way rather than another. Knowing that 3 x 5 = 30/2 doesn’t motivate us one way or another. And knowing facts about the world might well tell us what exists, and how to achieve what we want. Knowing such things might *direct* our existing desires in one way or another. But how could it make us want anything in the first place? What could knowing that there is food in the kitchen lead me to do anything without some desire (to eat, to cook…) to act upon? So judgements of reason cannot motivate action.

## Discussion

One way of escaping Hume’s argument is to claim that (P1) – that moral judgements can motivate actions – is false. To do good actions, we have to have the *desire* to be good as well. If moral judgements *don’t* motivate us on their own, then this argument gives us no reason to believe that moral judgements aren’t judgements of reason.

On this view, to say ‘Murder is wrong’ is simply to describe murder. Strictly speaking, someone could hold this belief, but not care about what is wrong, and so be quite willing to murder if it suited them (a sociopath perhaps?). If we think, like Hume, that moral judgments are motivating, this is because people almost always *do* care about morality, and want to do what is right. So making claims about what is right or wrong is something that is relevant to what they do. But the moral judgment alone doesn’t motivate them.

A second response would be to deny (P2) and argue that there are some judgments of reason that can motivate us. However, realism claims that moral judgments are matters of fact – they ascribe (natural or non-natural) properties to actions, states of affairs or people. It remains difficult to understand how such judgments could be motivating.

# Hume’s is-ought gap

Hume presents a second argument against cognitivism, and so moral realism, by drawing a famous distinction between sentences that talk about what *is* the case (judgements of reason) and moral judgements, which talk about what *ought* to be the case. What is the relation between what is and what ought to be? How, for instance, do we get from the fact that some action will cause pain to the claim that we ought not to do it? The connection is not obvious.

How is this an objection to cognitivism? Hume is commenting on how moral arguments work. Suppose I say ‘Eating meat causes animal suffering. Therefore, you shouldn’t eat meat.’ According to cognitivism, the conclusion states a truth, and this truth is inferred from the premise. But how is this a rational inference, Hume asks? The premise tells me how the world is; the conclusion tells me how the world ought to be. But I can’t infer one from the other. There is a ‘gap’ between what is and what ought to be, so that we can’t reason from one to the other. If moral judgments were true or false, we *would* be able to infer them from other truth claims, such as matters of fact. But we can’t. This is a reason to think that moral judgments don’t make truth claims, and so cognitivism is false.

# The argument from Hume’s fork

Hume’s fork is the claim that we can have knowledge of just two sorts of claim: relations of ideas or matters of fact. Relations of ideas can be discovered just by thinking, by recognising the truth of an analytic proposition or by deductive reasoning. To deny a relation of ideas is to contradict oneself. Matters of fact are claims about what exists, and they are established by sense experience and causal inference.

We should add that Hume understands the faculty of knowledge as ‘reason’. So because there are just two types of knowledge, reason makes just two types of judgment – judgments about relations of ideas and judgments about matters of fact. With this in place, Hume presents the following argument:

P1. There are only two types of judgements of reason, relations of ideas and matters of fact.

P2. Moral judgements are not relations of ideas.

P3. Moral judgements are not matters of fact.

C1. Therefore, moral judgements are not judgements of reason.

This raises an issue for cognitivism. If moral judgments are not judgments of reason, then according to Hume’s fork, we cannot have any knowledge of them. This is because, Hume goes on to argue, moral judgments are neither true nor false, but function in some other way (non-cognitivism).

Should we accept the argument? The arguments for, and against, (P1) are arguments over whether empiricism about knowledge is true. We won’t review those arguments here.

What of (P2)? Hume presents two arguments for this. First, relations of ideas are supposed to be certain – to deny them is a self-contradiction. But moral claims aren’t conceptual truths, like ‘black is the opposite of white’, or truths of logic or mathematics, like 3 x5 = 30/2. So what relations of ideas are moral claims supposed to be? Second, there is no relation of ideas that applies just to morality. Any relation that describes moral or immoral actions also applies to physical objects, but these aren’t moral or immoral. Take murder, for example, which involves one thing killing another. A plant can kill another plant. There is nothing in the idea of ‘killing’ that gives us moral wrongness.

We can object that murder is not simply killing. It is wilful, premeditated killing. But, Hume responds, this just means that the action has a different *cause*. But the relation between cause and effect, that one thing brings about the death of another, which we describe as ‘killing’, remains the same. That some event has a particular cause is a matter of fact, not a relation of ideas. It is up to the person who wants to claim that moral judgements are relations of ideas to show what relations of ideas they are, and how they are unique to morality.

How about (P3)? If we claim that moral judgements are a matter of fact, we must identify *which* fact. But, Hume says,

Take any action that is agreed to be vicious—wilful murder, for instance. Examine it in all lights, and see if you can find the matter of fact… that you call ‘vice’. However you look at it, all you’ll find are certain passions, motives, volitions, and thoughts; those are the only matters of fact in the case. The vice entirely escapes you as long as you focus on the object…

Let’s allow that the death was caused by an act of will. How is that the fact that it is wrong? We cannot, through empirical investigation, find the property of ‘moral wrongness’. The judgment that murder is wrong doesn’t state an empirical fact.

The conclusion of Hume’s argument follows from his premises, i.e. the argument is valid. So if Hume successfully defends each premise, (P1), (P2) and (P3), then he has shown that moral judgments are not judgments of reason, i.e. they are neither true nor false and so cognitivism – and therefore moral realism – is false.

## Discussion

To answer the objection, the moral realist needs to either deny Hume’s fork or meet Hume’s challenge of identifying some relation of ideas or some matter of fact that constitute moral judgments.

Non-naturalists will likely deny Hume’s fork, arguing that moral intuitions are not relations of ideas in Hume’s sense because they are not analytic truths. They are synthetic propositions that are self-evident and so aren’t established the way that other, empirical matters of fact are. But Hume’s objection puts pressure on intuitionists to say more about how this is possible.

Reductive naturalists will argue that moral judgments are matters of fact. However, it will take philosophical reasoning to show *which* matters of fact they are.

# A. J. Ayer’s verification principle

In the 1930s, a school of philosophy arose called logical positivism, concerned with the foundations of knowledge. It developed a criterion for when a statement is meaningful, called the principle of verifiability, also known as the verification principle. On A. J. Ayer’s version, presented in *Language, Truth and Logic*, the verification principle says that a statement only has meaning if it is either analytic or empirically verifiable.

A statement is analytic if it is true or false in virtue of the meanings of the words. For example, ‘Bachelors are unmarried’ is analytic and true; ‘Squares have three sides’ is analytic and false. A statement is empirically verifiable if empirical evidence would go towards establishing that the statement is true or false. For example, if I say ‘The moon is made of green cheese’, we can check this by scientific investigation. If I say ‘The universe has 600 trillion planets’, we can’t check this by scientific investigation in practice, but we can do so *in principle*. We know how to show whether it is true or false, so it is ‘verifiable’ even though we can’t actually verify it. Furthermore, we don’t need to be able to *prove* that an empirical claim is true or false. For empirical verification, it is enough for empirical evidence to raise or reduce the probability that a statement is true.

The principle can be understood as a development of Hume’s fork. However, while Hume’s ‘fork’ provides an account of what we can *know*, the verification principle is an account of what statements have *meaning*. But the verification principle defines meaning in terms of *how we can know* whether a statement is true or false. Unless there is some way of showing, at least in principle, that a statement is true or false, then it doesn’t really say anything, it doesn’t make a meaningful claim. The verification principle claims that the only alternative to knowing something analytically is to use empirical experience. So like Hume’s fork, it defends a form of empiricism.

Ayer applies the principle to ethical language. Moral judgments, such ‘murder is wrong’ or ‘pleasure is good’, are not analytically true. This seems clear. But, he argues, empirical investigation can’t show them to be true (or false) either. He agrees with Moore that ‘*X* is wrong’ cannot mean ‘*X* would cause unhappiness’ (or any other proposition substituting a natural property for ‘wrong’). The open question argument shows that it is never a contradiction to say ‘*X* would cause unhappiness, but it is right to do it nonetheless’. So Ayer agrees that moral naturalism is wrong. We can show that murder causes grief and pain, or that it is often done out of anger. But we cannot demonstrate, in the same way, that it is wrong. We can show that people pursue pleasure, but we can’t show that it is good. And so we can’t use empirical experience to discover what is right or wrong.

Non-naturalists, such as G E Moore, believed that moral judgements are about non-natural properties. While they are neither analytic nor empirically verifiable, they are nevertheless true or false. Ayer rejects this, and argues that we can’t establish the truth or falsity of a moral claim by appealing to ‘intuition’, as Moore claims, unless we are able to provide some criterion for deciding between conflicting intuitions. Given the verification principle, only an empirical criterion will do. But there is no empirical test that will establish which intuition is correct and which is incorrect. (If there were, then moral properties would be natural properties again.) And so, Ayer concludes, moral judgements don’t state truths or falsehoods at all and are therefore not genuinely meaningful.

## Discussion

As the last sentence shows, Ayer’s objection is to cognitivism. Moral realism fails because it assumes that cognitivism is true. The objection depends on the verification principle. But the principle faces a famous objection. According to the verification principle, the principle itself is meaningless. The claim that ‘a statement only has meaning if it is analytic or can be verified empirically’ is not analytic and cannot be verified empirically. But if the principle of verification is meaningless, then what it claims cannot be true. So if the principle is true, it is meaningless, and so not true. Obviously, if it is false, it is false. Either way it is not true. Therefore, it does not give us any reason to believe that ethical language is meaningless.

Ayer claims that the principle is intended as *a definition*, not an empirical hypothesis about meaning. In other words, it is intended to reflect and clarify our understanding of ‘meaningful’ uses of words. Ayer accepts that the principle isn’t obviously an accurate criterion of ‘literal meaning’, but that is why he provides arguments in specific cases, such as ethical language, which support it.

But in that case, the verification principle is only as convincing as the arguments that are intended to show that it is the right definition of ‘meaningful’. If we do not find the arguments convincing, the principle provides no independent support. Ayer accepts Moore’s rejection of naturalism. This still leaves Ayer’s challenge to Moore’s intuitionism hanging: can intuitionism – or any form of realism – provide an account of how we can decide between conflicting intuitions?

# Mackie’s argument from relativity

Mackie’s first argument is from relativity. In his book *Ethics*, Mackie starts from the common observation that, as a matter of fact, moral codes differ from one society to the next, i.e. there is relativity of morality to societies. According to one society, slavery is permissible under certain conditions; according to another, it is never permissible; or again, female circumcision is right v. it is wrong; or all people should be treated as equals v. people should be treated according to their caste. This claim, which we may call ‘descriptive relativism’, is a factual one, and one that certainly seems correct.

We can use this fact, Mackie argues, to show that there are no *objective* moral truths, but the argument is indirect. The mere fact that there is moral disagreement between societies doesn’t by itself show that moral realism is wrong. Societies have also disagreed on empirical matters of fact, e.g. some have thought that the Earth is flat, others that it is round; or again, that some people can magically move objects by thought v. there is no magic; and so on. Even scientific theories have disagreed, e.g. Newtonian ideas of space, time and motion v. Einstein’s theory of relativity. But in all these cases, we should be happy to say that there are objective truths, and some people just made mistakes. So disagreement over a claim doesn’t show that there is no objective truth. So the mere fact that societies have disagreed over morality still leaves open the possibility that there is objective moral truths, but some societies have held mistaken beliefs about what is morally good and right.

The argument against moral realism comes when we consider how we should understand and explain the moral disagreements between societies. The realist must argue that different societies, with their different ethical values and practices are all trying to get at the truth about ethics. The relativist argues that this is implausible, and we should understand the morality of a society as a reflection of its way of life.

With scientific disagreements, the best explanation is that different societies don’t have sufficient evidence to discover the truth. It is perfectly plausible to think that there is just one way the world is, empirically speaking, but it is not always easy to discover how that is. Our empirical beliefs are caused by, and change in response to, discoveries of what is true about the world. In contrast, says Mackie, the idea that two societies which disagree are both trying to find ‘the truth’ about ethics doesn’t sit well with an understanding of the history of societies and how ethical practices develop. It is far more plausible to say that different ways of life have given rise to different moral beliefs than to argue that societies’ different moral beliefs result from very inadequate or badly distorted perceptions of the one moral reality. There are different ways that human beings live, and they have developed different conventions about how to live, and these conventions are reflected in their moral judgments.

Mackie goes on to note that the realist can respond that there are *general ethical principles* that different societies share. For example, most societies have prohibitions on killing, lying, and theft, and encourage care of the weak. If disagreement supports the view that there is no objective moral truth, then agreement supports the view that there is. Different ethical practices reflect the different particular conditions in which different societies are situated, but not different ethical principles. This explains why societies disagree and recognises how and why different ethical practices develop without giving up on moral realism.

Mackie argues that this response to the argument from relativity is weak. At best, it shows only that the fundamental principles of morality are objective. Other moral judgments are relative to particular circumstances, so a judgment that, say, ‘stealing is wrong’ is true in some societies but could be false in others. Although this is the kind of thing that utilitarianism might defend, it does not reflect how most people understand morality. People hold their moral judgments not on the basis of general principles, but because something about the act arouses their disapproval. They have an ‘intuition’ that it is wrong. *Which* acts arouse people’s disapproval differs from one society to another, so we cannot argue that these moral judgments are objective.

## Discussion

There are several responses that the moral realist can make to Mackie’s last argument, that the response from realism to the argument from relativity is weak.

One response is to say that moral realism isn’t trying to describe how most people think about morality, it is trying to give the correct metaethical theory. For example, if utilitarianism is the correct normative ethical theory, then perhaps it is simply true that the only objective moral fact is given by the principle of utility, because there are just two moral properties: good, which is happiness, and right, which is maximising happiness. Nothing that Mackie has said shows otherwise.

On this point, we should note that the moral non-naturalist G E Moore would say exactly this. Moore argued that we reach moral knowledge by ‘intuition’. But when he talks about moral intuitions, he doesn’t mean people’s gut reactions, as Mackie seems to think. He means rational, ‘self-evident’ propositions, i.e. judgments that we can understand to be true when we consider them carefully. Moore went on to argue that we can only have intuitions in this sense about fundamental moral principles, and that our intuitions support utilitarianism! If Mackie is trying to attack intuitionism, he completely misunderstands it here.

A second response is to object to Mackie’s claim that any moral judgments that are relative to the circumstances of a society are not objective. This misunderstands the nature of moral reality, we could say, as well as the nature of truth. For example, some plants grow in hot countries but not in cold countries. So ‘Chilli plants will grow well’ is a relative truth – it is true in one country but not another. But this doesn’t make it any less *objective*. Whether a chilli plant will grow well in the country you are in is a mind-independent fact. We can even turn the relative truth into a universal truth by stating the conditions that apply, e.g. ‘Chilli plants will grow well in hot countries’. Moral realists can say the same about moral judgments. Some ethical practices will be permissible in some circumstances but not in others. Whether a moral judgment is true will depend on whether the practice is actually morally right or wrong in those circumstances. If someone’s ‘intuition’ is that ‘stealing is wrong’ and they live in conditions in which stealing is wrong, then their intuition is objectively true. One person can think ‘Chilli plants grow well’ and another, living in different conditions, can think ‘Chilli plants don’t grow well’, and they can both be objectively correct, given the conditions they live in. So two people, living in different conditions, can have conflicting intuitions about stealing and both be objectively correct, given the conditions they live under. They only make a mistake if they think ‘stealing is always wrong, in every society’ (and this is not true). And ‘Stealing is wrong under conditions C’ (if we can spell out the conditions) is not a relative truth at all, but a universal one.

# Mackie’s arguments from queerness

Mackie presents a second argument against moral realism, which he calls an argument from ‘queerness’. The oddity of moral properties and how we would know about them if they did makes it implausible to think that there are any moral properties. The argument has two aspects, metaphysical and epistemological.

## Metaphysical queerness

If there were moral properties, Mackie argues, they would have to be very different from anything else in the universe. His argument for this claim rests on the connection between morality and motivation. Moral judgements motivate us – we avoid actions we believe are wrong and try to do actions that are right. But that means, if there were moral properties, simply *knowing* what is good or bad, right or wrong, would be enough to motivate us to act in certain ways. For this to be true, ‘goodness’, say, would have to have ‘to-be-pursuedness’ built into it.

If this is a confusing idea, that’s Mackie’s point. How could an objective property motivate us in this way? How could there be some direct, immediate relation between some fact of the world and our desires? Just to know something true about the way the world is doesn’t entail being motivated to do anything about it. As we might say, the direction of fit is wrong.

We may add that, clearly, moral properties cannot be natural properties discovered by sense experience and science. None of the properties we discover this way are intrinsically motivating. So if there are moral properties, they must be non-natural properties. What Mackie’s argument is supposed to show is how peculiar such non-natural properties would have to be.

## Epistemological queerness

Suppose there were moral properties. If some actions, such as an act of courage, have the property of being objectively right; or again, if some states of affairs, such as being in pain or cowardice, have the property of being objectively bad – how could we know? Intuitionism, Mackie claims, says no more than that we have some special faculty – but this is a terrible answer that doesn’t explain how we have this knowledge at all. If we think of our usual ways of knowing about the world – sense perception, introspection, hypothetical reasoning, even conceptual analysis – none of these can explain knowledge of morality. To say that we know moral judgements to be true or false ‘by intuition’ is only to say that we don’t know them in any of the usual ways. The theory doesn’t give us any real answer as to *how* we know truths by intuition.

The non-naturalist might well reply that it is not only knowledge of morality that faces this objection. We can’t explain our knowledge of mathematics, necessary truths, the existence and nature of substance, space or causation in any of these ways either. Here, the non-naturalist is appealing to rationalist arguments about the scope of a priori knowledge. Is our knowledge of moral properties any more puzzling than our knowledge of these other things?

Mackie accepts the point: either empiricism can account for knowledge in these areas, or they all face the objection that they appeal to something ‘non-natural’. To a significant extent, then, Mackie’s argument depends on empiricism, rather than rationalism, being the correct account of our knowledge.

But Mackie presses the argument from epistemological queerness by asking what the connection between natural properties and moral properties is. For instance, we commonly say things like ‘that’s wrong because it is cruel’. If we take cruel to mean ‘causing pain for fun’, then cruelty is a natural property. It is a psychological fact that something causes pain, and another psychological fact that someone’s motive is taking pleasure in doing this. But what is the relation between these facts and the ‘fact’ that acting in this way is wrong? How can we establish whether it is wrong or not? It isn’t an analytic truth, and we can’t deduce it. Intuitionism fails to tell us how morality is related to anything else, how natural facts contribute to moral thinking. This makes it even more puzzling how we could come to know about moral properties.

## Discussion

Mackie’s argument from queerness depends on his understanding of what moral realism claims. In particular, he takes moral realism to be committed to the idea that moral properties are mind-independent and part of reality. Both these ideas need careful thought.

‘Reality’ here can’t mean simply the world as physics describes it – space, time, matter and perhaps causal relations between them. But obviously, physics won’t tell us right from wrong. But why should we think that all reality is like physical reality? Moral properties, if they exist, aren’t going to be like physical properties. Even reductive naturalists think the most likely natural properties to be moral properties are psychological properties.

Are psychological states ‘part of reality’? They certainly exist – whether one is happy or not is a psychological fact. In one sense, it is not a mind-independent fact, because it is a fact about a mind. In another sense, we can argue that it is a mind-independent fact, because whether you are happy or not is true or false independent of what anyone thinks. Anyone can make a mistake about whether or not you are happy, even you (you might think you are happy when, really, if you were completely honest with yourself, you’d realise you are not)! Perhaps this is controversial. So let’s talk about eudaimonia instead. Whether someone is eudaimon is, according to Aristotle, objectively true or false, but it is a fact about someone’s life, including their mind. There are lots of facts that are about human beings and their activities, e.g. not just psychological facts, such as whether someone is in love, but also cultural facts, such as facts about whether a piece of music is baroque or classical. But they are still facts, because they are independent of our judgments and made true by the way the world is, in this case, the human world.

Moral realism claims that moral judgments are mind-independent in the sense that whether a moral claim is true does not depend on whether we think that it is true. It doesn’t have to claim that moral judgments are not about minds. If moral facts are facts about our minds, perhaps they are not all that ‘queer’ after all.

Reductive naturalism, which claims that moral properties are, in fact, natural properties, argues that we can make the case even more strongly. Consider utilitarianism as an example. John Stuart Mill argues that our experience *does* give us evidence of what is good. What is good is what is desirable, and the best evidence for what is desirable is what people generally desire. Once we recognise this, there is no particular epistemological difficulty in discovering moral properties. Furthermore, if we say that goodness *is* happiness, then there is no metaphysical queerness about goodness either. It is simply a natural psychological property. And yet it is a motivating one. We desire happiness and are motivated to pursue it. Saying that something is good is to say that it produces happiness, so it is no puzzle how moral properties and moral judgments motivate us.

These responses are driven by (reductive or non-reductive) moral naturalism. We can show how moral properties are not queer by seeing how they fit with our understanding of human life. But Mackie’s arguments from queerness originally targeted non-naturalism. It is harder to see how non-naturalist theories such as Moore’s intuitionism can respond.

# Mackie’s Error theory

Mackie agrees with moral realists that we understand moral judgements to be cognitive. The way we use ethical language is to make objective claims about a moral reality. Moral judgments express beliefs about mind-independent moral properties, and these beliefs can be true or false. This is how moral language functions. But, he argues, there is no such moral reality. And so he argues that all moral judgements are false. ‘Murder is wrong’ is false, because the property of being wrong does not exist. ‘Murder is right’ is false, because the property of being right does not exist. There are no moral properties. Ethical language rests on a mistake.

Moral realism claims that there are moral properties, and these are objective – ‘mind-independent’ – because whether some action or state of affairs is good or bad, right or wrong, is independent of whether people believe that it has this property. But, Mackie argues, there are no mind-independent moral properties. This is the mistake that we make. Mackie is an ‘error theorist’.

Error theory is a form of moral anti-realism. It accepts cognitivism – moral language asserts claims about the world which are intended to be true. But error theory rejects moral realism: there are no mind-independent moral properties to make moral judgments true. Here is an analogy. Some people really believe in fairies. They don’t think that when we are talking about fairies, we are using language ‘fictionally’ (like when we talk about Sherlock Holmes). Imagine that everyone believed in fairies in this way. An ‘error theory’ of fairies would say that while talk of fairies is cognitivist, there are no fairies. It is not true that fairies have wings, because there are no fairies. It is not true that fairies don’t have wings, because there are no fairies. All claims about fairies are false, because there are no fairies.

The way Mackie phrases his argument is to say that there are no ‘objective’ moral values. A claim is objective, according to Mackie, if:

1. It can be something we know.
2. It can be true or false.
3. Its truth is independent of what we want or choose.
4. It is about something mind-independent.
5. It is about something that is part of reality

His claim is that moral judgments cannot be objective in any of these senses, because there are no moral properties.

## Discussion

We looked at Mackie’s defence of the claim that there are no objective moral properties when discussing his arguments from relativity and queerness. We also looked at responses to those arguments from moral realism. We won’t repeat those discussions here, but it is worth briefly revisiting one of them.

Moral realists agree with much of Mackie’s idea of objectivity. Moral realism wants to defend the claims that moral judgments can be true, and whether they are true does not depend on whether we want them to be so. It says that we can know some true moral judgments and that moral properties are real, so they are part of reality. However, in the discussion of Mackie’s arguments from queerness, we saw that the claim that moral judgments are mind-independent was ambiguous. Moral realists claim that moral judgments are mind-independent in the sense that they are true or false whether or not we *think* they are true or false. But they may not be mind-independent in the sense of being about something other than minds. It may be that moral judgments are about psychological properties, such as happiness. One response to Mackie’s error theory, then, is to say that he has misunderstood what it is for a moral property to be objective.

# Non-cognitivism and moral anti-realism

Non-cognitivist theories of ethics claim that ethical language does not try to describe the world and cannot be true or false. Moral judgements do not express beliefs, but some other, non-cognitive mental state. Different non-cognitivist theories disagree on exactly what moral judgments express, but they agree that moral language does not function to state facts.

EMOTIVISM

Non-cognitivist theories are anti-realist. Since moral judgments do not describe the world and are neither true nor false, then there are no mind-independent moral properties that would make moral judgments true or false. For example, to say ‘racism is wrong’ is not to claim that racism has any kind of property. It is, instead, according to emotivism, to express disapproval of racism.

# Emotivism and subjectivism

In one sense, emotivism claims that morality is ‘subjective’. However, there is an important distinction between emotivism and the theory that is called ‘subjectivism’. Subjectivism claims that moral judgements *assert* or *report* approval or disapproval, and there is a difference between expressing disapproval and asserting it. We can understand this better by looking at what subjectivism claims, and then contrasting it with emotivism.

One form of subjectivism claims that to say ‘*X* is wrong’ is simply to say that ‘X is generally disapproved of’. But this can’t be right, because it is not a contradiction to say ‘Most people approve of *X*, but *X* is wrong nonetheless’. For example, racism has been very common historically. We may argue that ‘racism is wrong’ even while acknowledging that most people approved of it.

A second form of subjectivism, ‘speaker subjectivism’, claims that the meaning of ‘*X* is wrong’ is something like ‘I disapprove of *X*’ or again ‘I think *X* is wrong’. This is a (psychological) fact about oneself, so the statement can be true or false. Speaker subjectivism, therefore, is an unusual form of reductive moral naturalism: the facts that make moral judgements true are facts about the individual speaker’s mind.

Speaker subjectivism entails that we cannot make mistakes about what is right or wrong. If I say ‘Murder is right’, I am simply stating ‘I approve of murder’. If I am sincere, then I do approve of murder, and so murder is, indeed, right (‘for me’, we might say). But, we can object, we naturally think that people *can* make mistakes about morality. Speaker subjectivism makes no sense of deliberation, trying to *figure out* what is right or wrong. Why should I bother to deliberate? *Whatever* I come to feel will be right!

By contrast, emotivism claims that moral judgements do not express *any* kind of truth or falsehood, because they are not cognitive. Where subjectivism is a form of cognitivism, emotivism is a form of non-cognitivism.

This enables emotivism to explain, and respond to, the objections to subjectivism above. To say that ‘most people approve of racism’ does not contradict ‘racism is wrong’, because ‘racism is wrong’ doesn’t state something true or false. It doesn’t *state* anything at all. Instead, it expresses the speaker’s disapproval of racism. And we cannot be infallible in the sense of getting the answer right; there are no moral truths.

# Why emotivism?

In *Language, Truth and Logic*, A J Ayer argues that moral judgments do not make claims that are true or false as they are neither analytic nor can they be verified empirically. But, according to his ‘verification principle’, these are the only two ways in which we can meaningfully make claims about what is true or false. So he concludes that moral language does not make meaningful claims about what is true or false. Instead, he argues, moral judgments express feelings:

If I say to someone, “You acted wrongly in stealing that money” … I am simply evincing my moral disapproval of it. It is as if I had said, “You stole that money,” in a peculiar tone of horror’. Through expressing our feelings, it also aims to arouse feelings in others, and so get them to act in certain ways.

Ayer’s defence of emotivism depends on the verification principle. However, defending the central claim that moral judgments express emotions, rather than state facts, does not depend on the principle of verification. For example, in *Facts and Values*, Charles Stevenson argues that moral words have emotive meanings, which are not descriptive. The central ethical terms – ‘right’, ‘wrong’, ‘good’ and ‘bad’ – only have emotive meanings, of expressing approval or disapproval. But many moral terms (‘steal’, ‘honesty’, ‘respect’) have both descriptive and emotive meanings. To be told that someone is ‘honest’ is to learn something about them. For instance, they can’t be honest while lying frequently! And whether someone lies frequently is a matter of fact. But the term ‘honest’ isn’t just a description; it also has an emotive meaning of approval.

The emotive meaning of moral judgments is related to their use. The purpose of moral judgements is not to state facts, but to influence how we behave through expressions of approval and disapproval. ‘*X* is good’, and other moral judgements, are used both to express the speaker’s attitudes and to influence the attitudes of other people. Moral terms are ‘dynamic’, and the main purpose of making moral judgements is to influence other people’s emotions and behaviour.

One advantage of emotivism is that it easily explains how and why it is that moral judgements motivate us. If moral language were just descriptive, stating how things are, why would that get us to act in certain ways? We need to care. And what we care about is captured in our attitudes to the world. Emotivism connects caring, approving, disapproving, with the very meaning of ethical words.

# Emotivism and moral language

Is emotivism’s analysis of the meaning of moral language correct? We can argue that it is not.

First, being emotive and influencing people’s attitudes is something that lots of non-moral language does as well, e.g. advertising. So we will need to say more to distinguish morality from advertising.

Second, does moral language always function to influence others? We may express our moral attitudes to others who already agree with them or that we know to be indifferent to our views – so influencing their attitudes is not the purpose. But this doesn’t show that we aren’t expressing a moral judgement.

Third, moral language isn’t always particularly or necessarily emotive. The key moral terms ‘good’, ‘right’, ‘wrong’ and ‘bad’ may arouse emotions in others or express ours, but again, this depends on context. We do not think that it is always good to arouse emotions in others on moral issues, especially by using emotive language. Moral discussion can be, and sometimes should be, dispassionate.

But how strong are these objections? The purpose of moral language, says emotivism, is to influence what people do. Without this, we would have no moral language or judgements at all. However, that doesn’t mean that it always has to be used for this purpose. This is normal – many types of language can be used in ‘non-standard’ ways in different situations. For example, it is possible to use fact-stating language to insult someone; e.g. ‘You have a big nose’. That it is an emotive statement on this occasion doesn’t make the meaning of the sentence ‘emotive’ – it states a factual claim. Likewise, language which is standardly emotive can be deployed without the intention to arouse emotion or influence action. The objections don’t show that moral language isn’t ‘essentially’ emotive, only that it isn’t always emotive.

However, it is worth noting that cognitivist theories don’t face this objection in the same way. According to cognitivist theories, moral judgments are statements of fact. The meaning of moral judgments is given by what would make them true. So we can understand what ‘murder is wrong’ means by understanding what it is for murder to be wrong. Now, we can use the claim ‘murder is wrong’ to do other things, such as influence people’s behaviour or complain or express anger or…. But these effects don’t give us the meaning of moral judgments. The many uses of moral language don’t threaten cognitivist theories the way that they threaten emotivism.

# Emotivism on moral reasoning

If I say ‘abortion is wrong’ and you say ‘abortion is right’, according to emotivism, it seems that I am just expressing my disapproval of it and you are expressing your approval. I’m just saying ‘Boo! to abortion’ and you’re saying ‘Hurrah! for abortion’. I am also trying to influence your attitudes, and you are trying to influence mine. But we are not doing so rationally, or by appealing to facts about what is good or bad. Trying to influence people without reasoning is just a form of manipulation. Emotivism reduces moral argument to propaganda. While sometimes moral argument might take this form, we do usually take ourselves to be reasoning about what is right, not simply mouthing off.

We can put the point another way. If, as emotivism claims, moral judgements and arguments are about influencing people’s attitudes, then a good moral argument will be one that is effective. That is all. There is no other, e.g. rational, criterion by which we might judge that it is a good or bad argument. Whatever I appeal to, to make you change your mind, no matter how irrelevant or far-fetched, if it makes you change your mind, it is a good argument. This is highly unsatisfactory.

Ayer responds that moral arguments are not arguments over moral judgements, but over facts: ‘we do not attempt to show by our arguments that [the other person] has the “wrong” ethical feeling towards a situation whose nature he has correctly apprehended. What we attempt to show is that he is mistaken about the facts of the case.’ When arguing over animal rights, say, we are constantly drawing facts to each other’s attention. I point out how much animals suffer in factory farms. You point out how much more sophisticated human beings are than animals. I point out that it is unkind to kill animals for food. You respond that people are not motivated by unkindness, and indeed, farmers can be very kind to the animals when alive. And so on. But if we both agree on the facts, but still disagree morally, there is nothing left to discuss, says Ayer, no further argument can take place. Moral judgements always presuppose a system of values; but no arguments for these values can be given.

But there are two objections to this response from Ayer. First, if you and I disagree about a moral judgement, and moral judgements have no truth value, are we right to say that there is a ‘disagreement’ here at all? Isn’t a disagreement when you think some claim is true and I think it is false? If so, then moral disagreement is only possible if cognitivism is true, since only cognitivism says that moral claims can be true or false.

Second, emotivism does not give us an adequate account of deliberation. If you are unsure about whether something, lying say, is right or wrong, we can understand that you are trying to work out what your attitude towards lying should be. But why can’t you settle the question of whether lying is right or wrong by simply noting whatever attitude you already have towards it? If emotivism is right, it seems that thinking hard about the question is irrational.

We can put the point another way: emotivism doesn’t explain how someone can change their mind on a moral issue because of rational argument. First, they have one attitude, then they have another. But what reason do they have to change their mind?

PRESCRIPTIVISM

Prescriptivism is another anti-realist theory. Since moral judgments do not describe the world and are neither true nor false, then there are no mind-independent moral properties that would make moral judgments true or false. For example, to say ‘racism is wrong’ is not to claim that racism has any kind of property. It is, instead, according to prescriptivism, to recommend that we do not act in racist ways.

When I express a moral judgement, Hare says, I am prescribing what you ought to do. I am not trying to influence or persuade you, nor am I expressing my feelings. Whether, as a result, you act as I prescribe is a different matter. Simply saying you should do x isn’t an attempt to persuade you – that may require a lot of argument.

So what is it to ‘prescribe’ something? There are two types of prescriptive meaning, Hare claims. First, there are imperatives that tell someone to do something. Imperatives explicitly state what to do, e.g. ‘Shut the door’. Hare argues that some moral judgements work in a similar way. For example, ‘Eating meat is wrong’ entails the imperative ‘Do not eat meat’. How so? To accept the imperative, ‘Shut the door’ is to shut the door. To accept that eating meat is wrong is to not eat meat. So if you ask ‘should I eat meat?’, and I answer ‘eating meat is wrong’, then I have answered your question.

Second, there are value judgements. The most general value terms are ‘good’ and ‘bad’. We use the word ‘good’, says Hare, when we want to commend something to someone. This commendation, although it is not explicit about what to do in the way imperatives are, provides guidance for our choices. How so?

# Good

We can talk about good chocolate, good teachers and good people. In each case, we are saying the chocolate, teacher or person is praiseworthy in some way. This use of language is quite distinct from describing something. Suppose I say ‘That’s a good strawberry, because it is sweet and juicy’. If we think ‘good’ as applied to strawberries just means ‘sweet and juicy’, then all I have said is ‘That’s a sweet and juicy strawberry because it is sweet and juicy’. But this isn’t what I said. I commended the strawberry, I didn’t merely describe it.

Because there is a distinction between describing and commending, nothing about being honest (i.e. telling the truth: descriptive meaning) can make me commend honesty (telling the truth is how to behave: prescriptive). More generally, nothing about the facts can entail a moral judgement.

However, ‘good’ is not purely a term of praise. Whenever we call something good, in each case there is a set of standards that we are implicitly relying on. Good chocolate is rich in the taste of cocoa. Good teachers can explain new ideas clearly and create enthusiasm in their students. A good person – well, a good person is someone who is the way we should try to be as people. When we use ‘good’ to mean ‘morally good’, we are appealing to a set of standards that apply to someone as a person. However, because nothing about the facts entails a moral judgement, there are no facts that establish one set of moral standards as objectively correct. We have to adopt the standards; they are not part of the world, waiting for us to discover them.

The descriptive meaning of ‘good’ in any context comes from the set of standards that is being assumed. Its descriptive meaning picks up on the qualities that the something must have to be (a) good … (chocolate, teacher, person, whatever). Because ‘good’ is always used relative to a set of standards, it always has a descriptive meaning. If you know what the standard for a good teacher is, then you learn something factual about a teacher when I say ‘she’s a good teacher’.

This has an important implication: if we have two identical things, we cannot call one of them good and the other not good. Whenever we apply a standard in making a prescription, we are committed to making the same judgement of two things that match the standard in the same way. If I say this chocolate is good but that chocolate is not, I must think that there is some relevant difference between the two.

‘Good’ is used primarily to commend. For each type of thing that we describe as good, the standard is different, but in each case, we are commending it. However, we don’t always use ‘good’ to commend. In fact, any word that both commends and describes can be used just to describe and not commend. For example, we often use the word ‘honest’ to commend someone. But I can say ‘If you weren’t so honest, we could have got away with that!’ This is an expression of annoyance, not praise. Likewise, I can agree that a ‘good person’ is one who is honest, kind, just, etc. But I can still think that good people are not to be commended, because, as Woody Allen said, ‘Good people sleep better than bad people, but bad people enjoy the waking hours more’.

So, according to Hare, the main features of ‘good’ are these:

1. It is used to commend, to provide guidance for choosing what to do.
2. It assumes a set of standards, features in virtue of which something counts as ‘good’ or not.
3. Two identical things must both be good or not. To think otherwise is logically contradictory.

# Moral language

In moral language, ‘good’ refers, directly or indirectly, to being a good person. A good action, then, will be one that a good person does. Calling something or someone ‘morally good’ is intended to guide people’s choices. The standards for who counts as a ‘good person’ are moral standards. However, moral standards are adopted, rather than being true or false.

The same three features that apply to ‘good’, Hare argues, also apply to ‘ought’ and ‘right’.

1. We say ‘you ought to pay back the money’ (in a particular situation) or again ‘stealing is wrong’ (in general) to guide people’s choices and actions.
2. The standards that we are assuming in making these judgements relate to being a good person.
3. Two actions, in similar situations, must either both be right or not. If I think that it is wrong for you to steal from me, because it infringes my rights of ownership, then I must think that it is wrong for me to steal from you, because it infringes your rights of ownership – unless I can say that there is some relevant difference between the two cases. We must be willing to ‘universalise’ our moral judgements. Not to do so is logically contradictory.

# Prescriptivism and moral language

Is prescriptivism’s analysis of the meaning of moral language correct? We can argue that it is not. Moral language does not only prescribe, but has many other functions. Hare has in mind the situation in which someone asks what to do. But there are lots of other situations in which we use moral language – we can exhort or implore someone, we can confess, we can complain, and so on.

However, Hare can reply that prescriptivism says that it is essential to morality that it guides choices and actions. This isn’t to say that, on every occasion, a moral judgement is being made to offer such guidance to the listener. The important point is that in holding a particular moral judgement, e.g. ‘stealing is wrong’, I am committed to acting on it.

We should accept this point. But it doesn’t necessarily support prescriptivism. It is not only commending and commanding that make a link between language and action. Language that expresses desires and attitudes also makes such links. Suppose I say ‘I like apples’, but I never eat apples, refuse anything made from apples, etc. There is something inconsistent here. Likewise, I can say ‘I disapprove of stealing’, but steal myself and never comment on others’ thefts. Just by connecting ethical language to action, Hare hasn’t shown that ethical language must be prescriptive. It could just as well express what we want or our attitudes.

It is worth noting that cognitivist theories don’t face this objection in the same way. According to cognitivist theories, moral judgments are statements of fact. The meaning of moral judgments is given by what would make them true. So we can understand what ‘murder is wrong’ means by understanding what it is for murder to be wrong. Now, we can use the claim ‘murder is wrong’ to do other things, such as influence people’s behaviour or complain or express anger or…. But these effects don’t give us the meaning of moral judgments. The many uses of moral language don’t threaten cognitivist theories the way that they threaten prescriptivism.

# Prescriptivism on moral reasoning

One use of moral language is in moral reasoning – using premises to draw conclusions about what to do. If I say ‘abortion is wrong’ and you say ‘abortion is right’, according to prescriptivism, it seems that I am just prescribing that you and I should not abort while you are refusing the prescription. But are we doing so rationally, or by appealing to facts about what is good or bad?

Hare argues that prescriptivism can explain moral reasoning. First, we can ask about someone’s reasons for prescribing what they do. Second, morality involves consistency – moral judgements must be universalised. For example, Singer claims there is no relevant difference between the suffering of people and the suffering of animals. If we are going to say that causing the suffering of people is wrong, we are committed to saying the suffering of animals is wrong – unless we can find a relevant difference. Moral disagreements can be about the consistency in applying certain standards, and reason can help resolve this.

Third, we can infer prescriptions from other prescriptions. A famous argument against abortion says ‘Taking an innocent human life is wrong. Abortion is the taking of an innocent human life. Therefore abortion is wrong.’ This has the same logical force, Hare claims, if we rephrase it as imperatives: ‘Do not take innocent human life. Abortion is the taking of an innocent human life. Therefore, do not commit abortion.’ To reject or refuse the conclusion, we must reject or refuse at least one premise. And so our prescriptions are logically related to one another. So we can use reason to discuss these relations. Moral arguments are not only about the facts, but about moral judgments as well.

We can object that the only kind of rationality prescriptivism can recognise in moral arguments is consistency. In requiring us to universalise moral judgements, Hare’s theory is similar to Kantian ethics. However, Kant argues that the standards for a good person (the good will) are themselves set by reason, and are therefore objective. Hare rejects this. Neither the empirical facts nor reason entails that we must have certain standards rather than others. If I argue that racism is morally right, and equality is morally wrong, as long as I am prepared to universalise this claim, there is no objective ground on which to disagree with me. Suppose you say ‘But what if you were of a different race. Would you say you should be treated as inferior?’ I can reply ‘Yes.’ Now what?

Hare responds that to prescribe that one’s own interests be frustrated like this is irrational. And so his moral system will give us the Golden Rule of ‘Do unto others as you would have them do unto you’ – anything else would be inconsistent.

But we can press the objection in a different form. Suppose you say that we shouldn’t steal because stealing would make life very difficult. This is your reason for prescribing that we shouldn’t steal, implicitly appealing to the standard that the good person does not make life difficult. But Hare says that moral standards are not objectively correct. Suppose I do not adopt your standard – I have a different standard for ‘good’. Then not only do I reject your moral judgement that stealing is wrong, I also don’t accept your reason for this judgement as a moral reason. So, on Hare’s view, there are no reasons to do a particular action independent of what standards we adopt. And so moral rationality is no more than consistency. But this does not rule out very objectionable values.

OBJECTIONS TO MORAL ANTI-REALISM

# DOES moral anti-realism become moral nihilism?

Moral nihilism is the rejection of all moral values and principles. It is the view that nothing is of moral value, that we have no moral duties. Moral anti-realism claims that there are no mind-independent moral properties, no objective moral truths, and non-cognitivist forms of anti-realism claim that morality is an expression of our emotions or attitudes. We may object that if this is so, then really there are no moral values – we invent them. Really, we have no obligation to be moral, because we have no obligation to have certain emotions or adopt certain standards of value. If moral properties are not objective, if moral judgments are not objectively true, then why accept morality at all?

Moral anti-realists can argue that this is either an unfair simplification of their theories or a straightforward misunderstanding. Error theory allows that we can have subjective moral values, and all three theories will argue that living without moral values is itself a choice or expression of feeling, and one that moral people will disapprove of morally. The theory that moral values are a reflection of our feelings does not imply that we should stop having moral feelings. The emotivist may still show disapproval of anyone who advocates that morality doesn’t matter or is just a matter of taste. Similarly, the fact that we must adopt standards of value doesn’t show that we should stop making prescriptions, and we may prescribe that people live according to particular standards.

The moral nihilist can respond that this is unjustifiable. Can we really justify interfering with how other people behave – when they behave ‘immorally’ – just because their actions don’t accord with our feelings or choices? This seems very petty. But this isn’t the reason we are interfering, claims the anti-realist. It is not because it offends us, but because they are being racist or cruel or cowardly or whatever.

The difficulty here is that my taking racist discrimination as a good reason to prevent an action is itself an expression of my feelings or the standards on which I make prescriptions. For the anti-realist, moral disagreement is always eventually a stand-off between subjective points of view. For the moral realist, by contrast, that racist discrimination is a good reason to interfere with someone’s action is a moral fact, based on the moral properties of racism. The moral realist claims to have the backing of reality.

# hOW CAN WE ACCOUNT FOR Moral progress

If there is no moral reality, we can argue, then our moral views cannot become better or worse. Obviously, they have changed – people used to believe that slavery was morally acceptable and now they do not. But how can we say that this is progress if there is no objective moral truth? There are two responses moral anti-realists can give.

First, they can claim that there can be very real improvements in people’s moral views if they become more rational. This can happen in several different ways.

1. People may come to know certain facts that they didn’t know before. In the case of slavery, people believed many things about slaves that were not true (one popular false belief was that they were stupid). Moral progress here means basing one’s morality on the facts, not false beliefs.
2. People can become more consistent, more willing to universalise their principles. For example, Singer argues that if we were consistent in our feelings about preventing suffering, we would not eat meat. If he is right, then vegetarianism would be moral progress.
3. People can become more coherent in their moral judgements. Many of us have moral feelings that come into conflict with each other, e.g. over lying. Moral progress here would be a matter of working out the implications of our views, and changing what needed changing to make them coherent with each other.

Because people are ignorant, do not always think logically, and have not resolved the conflicts between their different feelings and conventions, there is plenty of room for moral progress. But moral progress just means becoming more rational in our moral thinking, not becoming more ‘correct’ in our moral judgements.

The second response moral anti-realists can give is this: if we disapprove of past moral codes and approve of our own moral code, then we will say that we have made moral progress. Society has moved from moral principles that were bad (i.e. principles we disapprove of) to moral principles that are good (i.e. principles we approve of). That is what moral progress is.

This response means that moral progress is relative to a particular moral point of view. Non-cognitivists will say that talk of moral progress is itself non-cognitive, an expression of someone’s moral attitudes rather than a claim that can be true or false. If two people disagree over whether we have made moral progress in the last 200 years, say, this disagreement should be understood in the same way that any moral disagreement is understood. There is no special problem about how to explain moral progress.