

Kant's Categorical Imperative¹

'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. 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. In this handout, 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? (We discuss why Kant thought this was the fundamental principle of morality in the handout 'Kant: the good will, duty and the Categorical Imperative'.)

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

¹ This handout is based on material from Lacewing, M. (2017) *Philosophy for AS and A Level: Epistemology and Moral Philosophy* (London: Routledge), Ch. 3, pp. 253-9

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 dentist 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.