

Kant's ethics

So far in our discussion of ethics we have been focusing on different versions of **consequentialism** - the view that one is morally obliged to pursue the course of action which, of the available alternatives, will produce the best outcome. Last time we focused on objections to consequentialism which turn, in part, on that view's indifference to the way in which the consequences of an action are brought about.

These objections might suggest that we should develop our ethical theory from a different starting point than the consequentialist; perhaps we should focus not on the consequences of our actions, but rather on **the actions themselves** in determining what we morally ought to do. This line of thought was developed by the author of one of the great non-consequentialist moral systems, Immanuel Kant.

In the selection from Kant's book, *The Groundwork of the Metaphysics of Morals*, which we read for today, Kant begins with this claim:

It is impossible to conceive anything at all in the world, or even out of it, which can be taken as good without qualification, except a **good will**.

The question which Kant then tries to answer is: what is it for a will to be good?

We have already seen the consequentialist's answer to this question: according to a consequentialist, for a will to be good is for it to aim at acting so as to produce the best possible state of affairs. In this sense, the consequentialist thinks that what is "good without qualification" are states of affairs; good wills are defined in terms of the intention to produce these good states of affairs.

According to Kant, this gets things exactly backwards:

This tells us what Kant denies: he denies that we can explain what makes a will good in terms of the consequences that will brings about. But then what **does** make a will a good will?

What makes a will good is its conformity with the moral law, which Kant called the **categorical imperative**.

A good will is not good because of what it effects or accomplishes—because of its fitness for attaining some proposed end: it is good through its willing alone—that is, good in itself.

if by its utmost effort it still accomplishes nothing, and only good will is left (not, admittedly, as a mere wish, but as the straining of every means so far as they are in our control); even then it would still shine like a jewel for its own sake as something which has its full value in itself.

According to Kant, this gets things exactly backwards:

This tells us what Kant denies: he denies that we can explain what makes a will good in terms of the consequences that will brings about. But then what **does** make a will a good will?

What makes a will good is its conformity with the moral law, which Kant called the **categorical imperative**.

A good will is not good because of what it effects or accomplishes—because of its fitness for attaining some proposed end: it is good through its willing alone—that is, good in itself.

if by its utmost effort it still accomplishes nothing, and only good will is left (not, admittedly, as a mere wish, but as the straining of every means so far as they are in our control); even then it would still shine like a jewel for its own sake as something which has its full value in itself.

there is an imperative which, without being based on, and conditioned by, any further purpose to be attained by a certain line of conduct, enjoins this conduct immediately. This imperative is **categorical**.

Here Kant distinguishes the moral law - the categorical imperative - from other rules of action, which he calls **hypothetical imperatives**. An example of a hypothetical imperative is: “Get something to drink, if you’re thirsty and don’t have any other pressing obligations.” This is a hypothetical imperative because it tells us what we should do, **given that certain other conditions are satisfied**. The categorical imperative is not like this: it, as Kant says, “enjoins the conduct immediately.” The categorical imperative tells us what we are morally obliged to do, period - no matter what.

This tells us about the status of the categorical imperative - that it tells us what we must do, no matter what - but what does the categorical imperative, itself, say?

Kant thought that there was exactly one moral rule, and that it can be stated as follows:

What makes a will good is its conformity with the moral law, which Kant called the **categorical imperative**.

Here Kant distinguishes the moral law - the categorical imperative - from other rules of action, which he calls **hypothetical imperatives**. An example of a hypothetical imperative is: "Get something to drink, if you're thirsty and don't have any other pressing obligations." This is a hypothetical imperative because it tells us what we should do, **given that certain other conditions are satisfied**. The categorical imperative is not like this: it, as Kant says, "enjoins the conduct immediately." The categorical imperative tells us what we are morally obliged to do, period - no matter what.

This tells us about the status of the categorical imperative - that it tells us what we must do, no matter what - but what does the categorical imperative, itself, say?

Kant thought that there was exactly one moral rule, and that it can be stated as follows:

There is therefore only a single categorical imperative and it is this: "*Act only on that maxim through which you can at the same time will that it should become a universal law.*"

Kant calls this the **formula of universal law**.

Your **maxim** is your reason for acting. The formula of universal law therefore says that you should only act for those reasons which have the following characteristic: you can act for that reason while at the same time willing that it be a universal law that **everyone** adopt that reason for acting.

The best way to understand what this means is by looking at Kant's discussion of an action which violates the formula of universal law.

2. Another finds himself driven to borrowing money because of need. He well knows that he will not be able to pay it back; but he sees too that he will get no loan unless he gives a firm promise to pay it back within a fixed time. He is inclined to make such a promise; but he has still enough conscience to ask "Is it not unlawful and contrary to duty to get out of difficulties in this way?" Supposing, however, he did resolve to do so, the maxim of his action would run thus: "Whenever I believe myself short of money, I will borrow money and promise to pay it back, though I know that this will never be done." Now this principle of self-love or personal advantage is perhaps quite compatible with my own entire future welfare; only there remains the question "Is it right?" I therefore transform the demand of self-love into a universal law and frame my question thus: "How would things stand if my maxim became a universal law?" I then see straight away that this maxim can never rank as a universal law and be self-consistent, but must necessarily contradict itself. For the universality of a law that every one believing himself to be in need may make any promise he pleases with the intention not to keep it would make promising, and the very purpose of promising, itself impossible, since no one would believe he was being promised anything, but would laugh at utterances of this kind as empty shams.

There is therefore only a single categorical imperative and it is this: *"Act only on that maxim through which you can at the same time will that it should become a universal law."*

Kant calls this the **formula of universal law**.

Your **maxim** is your reason for acting. The formula of universal law therefore says that you should only act for those reasons which have the following characteristic: you can act for that reason while at the same time willing that it be a universal law that **everyone** adopt that reason for acting.

The best way to understand what this means is by looking at Kant's discussion of an action which violates the formula of universal law.

Kant's line of reasoning here appears to be this: if I consider the maxim

Promise to get money whenever I need it with no intention of paying it back.

as a universal law, then I imagine a scenario in which everyone is constantly making false promises. But in this sort of scenario, the convention of promising would cease to exist: after all, no one would have any reason to lend money on the basis of promises if such promises are never kept. So in such a world it would be impossible to act on this maxim.

How could this line of reasoning be used to show that lying, in general, violates the formula of universal law?

Our discussion so far is already enough to bring out some important contrasts between Kant's ethics and the consequentialist ethical systems we have discussed.

There is therefore only a single categorical imperative and it is this: *“Act only on that maxim through which you can at the same time will that it should become a universal law.”*

Kant calls this the **formula of universal law**.

Your **maxim** is your reason for acting. The formula of universal law therefore says that you should only act for those reasons which have the following characteristic: you can act for that reason while at the same time willing that it be a universal law that **everyone** adopt that reason for acting.

The best way to understand what this means is by looking at Kant's discussion of an action which violates the formula of universal law.

Kant's line of reasoning here appears to be this: if I consider the maxim

Promise to get money whenever I need it with no intention of paying it back.

as a universal law, then I imagine a scenario in which everyone is constantly making false promises. But in this sort of scenario, the convention of promising would cease to exist: after all, no one would have any reason to lend money on the basis of promises if such promises are never kept. So in such a world it would be impossible to act on this maxim.

How could this line of reasoning be used to show that lying, in general, violates the formula of universal law?

Our discussion so far is already enough to bring out some important contrasts between Kant's ethics and the consequentialist ethical systems we have discussed.

First, Kant's ethics tells you, in the first instance, what morality **forbids** you from doing. But it does not tell you what you ought to do in every case; some actions might be morally praiseworthy even though not doing them would not be contrary to the Formula of Universal Law, and hence not morally forbidden. These actions are, therefore, neither morally required nor morally forbidden. For the consequentialist, on the other hand, one must always do what will bring about the best consequences: so (excluding ties) every action is either morally required or morally forbidden.

Second, according to the consequentialist, the rightness or wrongness of a particular action depends on which action, in these particular circumstances, would lead to the best outcome. According to Kant, by contrast, the rightness or wrongness of acting from a particular maxim just depends on the **type** of maxim that it is. If making false promises, or lying, is sometimes morally forbidden, then it is **always morally forbidden**.

This last point -- that the rightness or wrongness of an action just depends on the type of maxim from which one is acting, rather than on the consequences of this particular action -- explains why Kantian ethics yields quite different results about what we ought to do than the sorts of consequentialist views we discussed.

Our discussion so far is already enough to bring out some important contrasts between Kant's ethics and the consequentialist ethical systems we have discussed.

First, Kant's ethics tells you, in the first instance, what morality **forbids** you from doing. But it does not tell you what you ought to do in every case; some actions might be morally praiseworthy even though not doing them would not be contrary to the Formula of Universal Law, and hence not morally forbidden. These actions are, therefore, neither morally required nor morally forbidden. For the consequentialist, on the other hand, one must always do what will bring about the best consequences: so (excluding ties) every action is either morally required or morally forbidden.

Second, according to the consequentialist, the rightness or wrongness of a particular action depends on which action, in these particular circumstances, would lead to the best outcome. According to Kant, by contrast, the rightness or wrongness of acting from a particular maxim just depends on the **type** of maxim that it is. If making false promises, or lying, is sometimes morally forbidden, then it is **always morally forbidden**.

This last point -- that the rightness or wrongness of an action just depends on the type of maxim from which one is acting, rather than on the consequences of this particular action -- explains why Kantian ethics yields quite different results about what we ought to do than the sorts of consequentialist views we discussed.

Suppose, for example, that a judge knows that the defendant in a capital case is innocent, but also knows that not finding the defendant guilty and sentencing him to death will result in riots in which many will be killed. What would a consequentialist say about this sort of case? How about the Kantian?

In this sort of case, it might seem that the Kantian gets things right, and the consequentialist gets things wrong. But there are other cases where things might not seem so clear. Here is one such example:

You're living in Nazi Germany, and hiding a Jewish family in your basement. The authorities come to the door, and ask you whether you are hiding a Jewish family in your house. You know that they will believe you if you tell them that you are not; it is just a random check. What should you do?

What does the Kantian say about this sort of case? How about the consequentialist? What should we say?

Kant himself was well aware of this consequence of his theory, and he believed it to be correct. Thinking that one should lie to save someone's life is, for Kant, making a mistake about the nature of the moral law. It is not a hypothetical imperative, which tells you what you ought to do under certain conditions (such as those conditions in which it will lead to favorable outcomes) - it is a categorical imperative, which simply tells you what you must do, come what may.

One thing to think about is whether Kant's position on this sort of case is defensible. But there are other problems with the formula of universal law, quite apart from its consequences for particular moral dilemmas.

One thing to think about is whether Kant's position on this sort of case is defensible. But there are other problems with the formula of universal law, quite apart from its consequences for particular moral dilemmas.

There is therefore only a single categorical imperative and it is this: "*Act only on that maxim through which you can at the same time will that it should become a universal law.*"

One sort of problem arises from the fact that the formula of universal law seems to deliver quite different results depending on how we formulate our maxims in a given case. Consider the case of the Nazi at the door. We could formulate our maxim as any of the following:

- A. Lie whenever doing so would lead to a desired outcome.
- B. Lie whenever doing so would save someone from the Nazis.
- C. Lie whenever doing so would save someone from the Nazis, whenever so doing would never be discovered by the Nazis.

Maxim A fails the formula of universal law - and for analogous reasons, maxim B seems to as well. But how about maxim C? Would there be anything contradictory about acting on maxim C in a world in which everyone acted on maxim C?

This difference should be a bit worrying for the Kantian; it is not, after all, easy to see how one could even tell whether B or C is one's real reason for acting. And it is also odd that acting from maxim B should be morally forbidden, but not acting from maxim C.

These worries about identifying the relevant maxim are connected with a second worry about the formula of universal law: that it does not cover nearly enough cases to be the **single** moral law.

Consider, for example, the maxim governing the action of a man who abuses his wife. Suppose it is: "Physically abuse your wife whenever you feel like it." Is there any contradiction in imagining everyone acting on this maxim? Would it be impossible to act on this maxim in a world in which everyone did so? If not, then it seems to follow from the status of the formula of universal law as the single moral law that the man's actions are morally permissible. But this is surely a mistake.

The defender of the formula of universal law might reply by saying that we have incorrectly identified the man's maxim. Perhaps it instead should be: "Physically abuse anyone whenever you feel like it." Certainly it does not seem as though anyone would be rational to will that **this** maxim be universal law. But, even if a world in which this maxim was a universal law would be unpleasant, it does not seem that there is any **contradiction** in acting on this maxim in such a world; and, moreover, what tells us that this maxim, rather than the more specific one considered above, must be the man's maxim?

One thing to think about is whether Kant's position on this sort of case is defensible. But there are other problems with the formula of universal law, quite apart from its consequences for particular moral dilemmas.

There is therefore only a single categorical imperative and it is this: *“Act only on that maxim through which you can at the same time will that it should become a universal law.”*

The defender of the formula of universal law might reply by saying that we have incorrectly identified the man's maxim. Perhaps it instead should be: “Physically abuse anyone whenever you feel like it.” Certainly it does not seem as though anyone would be rational to will that **this** maxim be universal law. But, even if a world in which this maxim was a universal law would be unpleasant, it does not seem that there is any **contradiction** in acting on this maxim in such a world; and, moreover, what tells us that this maxim, rather than the more specific one considered above, must be the man's maxim?

However, the formula of universal law was not the only interpretation of the moral law Kant gave. He also thought that the categorical imperative could be stated as the following **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.

One might wonder how Kant could give these two formulations of the categorical imperative if he thought that there was just a single moral law. The answer is that Kant thought, roughly speaking, that the formula of universal law and the formula of humanity were just two ways of stating the same thing; that is, that they are two different ways of expressing a single moral law.

It is, to say the least, not easy to see why Kant thought this; the formula of universal law and the formula of humanity certain seem to say different things, even if they might deliver the same verdicts in many cases. But for now let's simply set aside the question of the relationship between these claims and ask instead: can the formula of humanity serve as the moral law?

Let's begin by asking: what does it mean to treat someone as an end vs. as a means?

This distinction is difficult to explicate in an uncontroversial way; but I think that it is also a distinction on which we have a clear intuitive grip. Think of the complaint that someone is simply **using** you. When we say this, we are saying that the person is not taking you into account; that he is treating you as a vehicle for his own ends, rather than as deserving respect and consideration in your own right. This is treating someone as a mere means rather than as an end in himself.

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.

Let's begin by asking: what does it mean to treat someone as an end vs. as a means?

This distinction is difficult to explicate in an uncontroversial way; but I think that it is also a distinction on which we have a clear intuitive grip. Think of the complaint that someone is simply **using** you. When we say this, we are saying that the person is not taking you into account; that he is treating you as a vehicle for his own ends, rather than as deserving respect and consideration in your own right. This is treating someone as a mere means rather than as an end in himself.

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

The formula of humanity has a powerful intuitive appeal, and seems to say the right thing about many of the difficult dilemmas we've discussed. What would the formula of humanity say about the case of the unwilling transplant? What about the case of pushing the man on the tracks to stop the trolley?

The formula of humanity is also uncompromising in much the way the formula of universal law is. Because it is a genuinely categorical imperative - one which says what you are morally required to do, no matter what the circumstances - it will often require actions which, from a consequentialist point of view, seem horrible. For example, what will the formula of humanity require in the case of the Nazi at the door?

What would the formula of humanity say about self-defense? Or shooting at the enemy in a war?

Or consider a variant of the trolley case, in which there are 1000 people on the tracks ahead, who can be saved by diverting the trolley to kill one. Can we really be morally required not to turn the trolley?

Many of the problems which arose for the consequentialist involve cases in which act-types which we are inclined to regard as morally wrong nevertheless bring about the best consequences - in those cases, the consequentialist seems committed to the incorrect judgement that we are morally obliged to perform the relevant action; and this looks good for the Kantian, who, it seems, correctly regards these actions as morally prohibited. But if we make the differences between the consequences more and more dramatic, to many it seems that it gets harder and harder to maintain the Kantian position.

Many of the problems which arose for the consequentialist involve cases in which act-types which we are inclined to regard as morally wrong nevertheless bring about the best consequences - in those cases, the consequentialist seems committed to the incorrect judgement that we are morally obliged to perform the relevant action; and this looks good for the Kantian, who, it seems, correctly regards these actions as morally prohibited. But if we make the differences between the consequences more and more dramatic, to many it seems that it gets harder and harder to maintain the Kantian position.

This leads to a final point. It is important to emphasize that, just as utilitarianism is but a single version of consequentialism, so Kant's moral theory is but a single version of a family of ethical views sometimes called **deontological** theories. Unlike consequentialists, deontologists judge the rightness or wrongness of acts not by goodness of consequences, but by accord with or violation of certain moral rules. One might endorse Kant's deontological approach to ethics without thinking that he correctly identified the moral rules - and without thinking that he was correct in thinking that there is exactly one moral rule.

Nor do consequentialist and deontological theories exhaust the field; an important tradition in ethics - the tradition of virtue ethics - holds that we can only understand moral rightness and wrongness in terms of an account of the virtues of character - an account, in the first instance, of what makes a good person, rather than a right action.

Nor do consequentialist and deontological theories exhaust the field; an important tradition in ethics - the tradition of virtue ethics - holds that we can only understand moral rightness and wrongness in terms of an account of the virtues of character - an account, in the first instance, of what makes a good **person**, rather than a right action.

And there are yet other approaches to ethics. Our brief survey has for this reason not been anything like a comprehensive survey of the range of ethical theories, though hopefully it has provided some understanding of the challenges that such theories face.