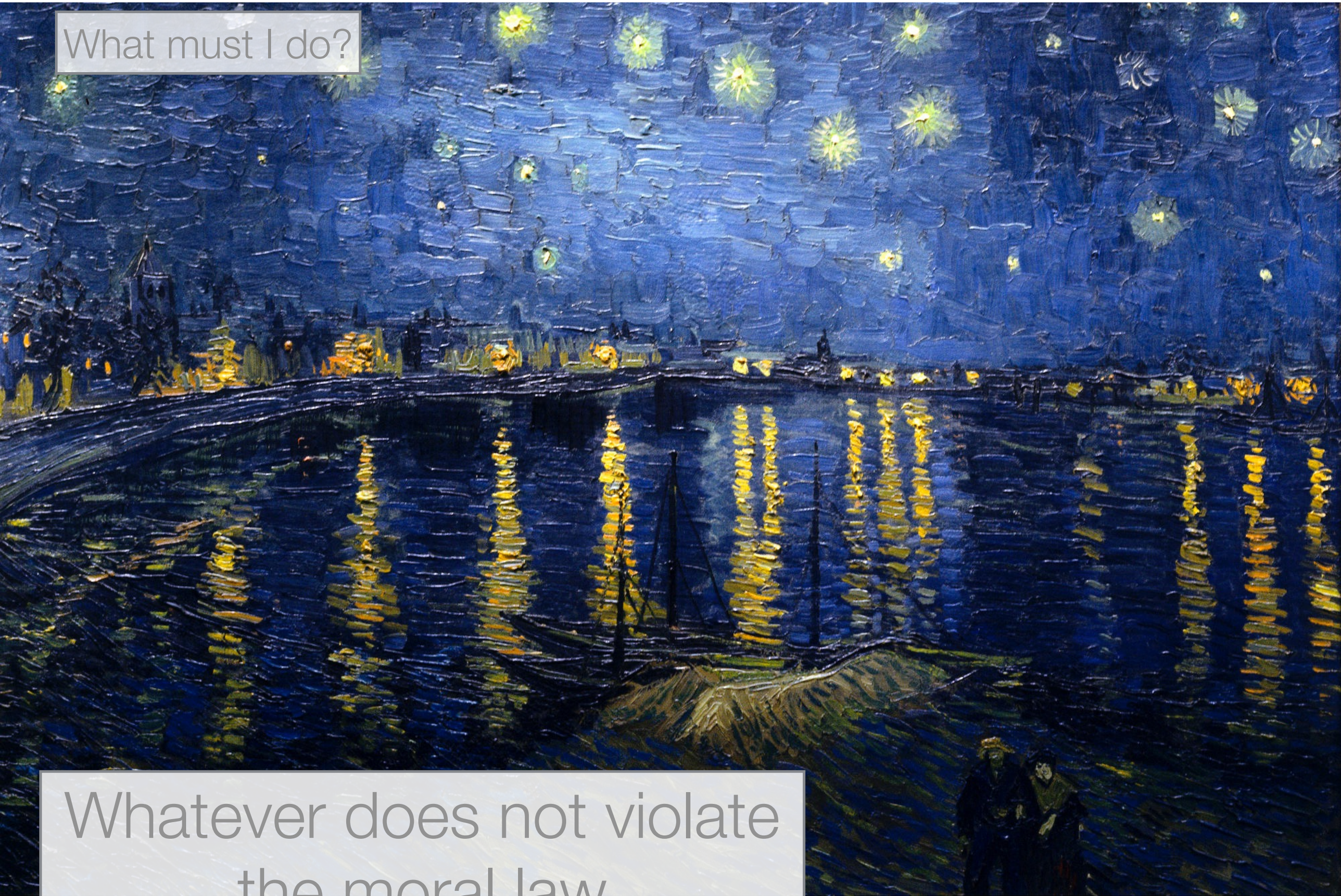
What must I do?



Whatever does not violate the moral law

Our question is: what makes some actions right, and others wrong? We've already seen the consequentialist's answer to this question. But there is another answer available.

This is the view that morality consists of a series of rules. Some examples might be:

Don't kill innocent people.

Don't demean people for being different than you.

Don't tell lies.

Always respect the wishes of other people.

On a rule-based approach to morality, behaving morally consists in not violating the rules. Might an approach like this help with some of the problems which, as we saw, consequentialism faces?

But, even if this sort of approach looks promising, it faces a basic question. For any list of moral rules, we can ask: Just why are those the moral rules? What makes them, rather than some other rules, the rules that we ought to live by?

We're going to talk about two answers to this question — two different versions of rule-based ethics.

The first answer to our question is a very old one. It says: those are the moral rules because they are the will of God. This approach to ethics is sometimes called divine command theory.

What one believes about ethics, on this sort of view, depends a lot on what one believes about God, and the will of God. A Catholic and a Muslim may both be divine command theorists, but might still disagree a lot about ethical questions — because they disagree about what God has commanded.

Today we will focus less on the choice between different views of what God has commanded, and instead will focus on a problem — first raised by Plato — which is a problem for any version of divine command theory.

The problem arises in a dialogue between Euthyphro and Socrates. Here is the way that Euthyphro states his view:

> EUTHYPHRO: Yes, I would indeed affirm that holiness is what the gods all love, and its opposite is what the gods all hate, unholiness.

Socrates responds by raising the following dilemma for this position:

> SOCRATES: We shall soon know better about that, my friend. Now think of this. Is what is holy holy because the gods approve it, or do they approve it because it is holy?

At first, Euthyphro is confused by the question. Socrates responds to his confusion with a series of examples, one of which uses the example of vision. Given that for any thing x, someone sees x if and only if x is seen, we can still ask: is x seen because someone sees x, or does someone see x because x is seen? The answer seems clear: it is the first. Something is seen because someone sees it, and not the other way around.

But then we can ask a parallel question about the moral law and what God commands. Socrates and Euthyphro agree that, for any action x,

God commands us to do x if and only if x is morally right.

But, even if we agree about this, we can still ask: does God command us to do x because x is morally right, or is x morally right because God commands us to do it?

It seems that there are two possible answers to this question:

(1) God commands us to do x because x is morally right.

(2) x is morally right because God commands us to do x.

Two things are pretty clear: the divine command theorist is committed to answer (2), and (1) and (2) are exclusive: they can't both be correct.

(1) God commands us to do x because x is morally right.

(2) x is morally right because God commands us to do x.

Socrates' problem for the divine command theorist takes the shape of an argument for (1):

SOCRATES: Then what are we to say about the holy, Euthyphro? According to your argument, is it not loved by all the gods?

EUTHYPHRO: Yes.

SOCRATES: Because it is holy, or for some other reason?

EUTHYPHRO: No, it is for that reason.

SOCRATES: And so it is because it is holy that it is loved; it is not holy because it is loved.

(1) God commands us to do x because x is morally right.

(2) x is morally right because God commands us to do x.

One can think of the argument as beginning with the thought that God must have some reason for issuing the commands that God does; otherwise, those commands would be completely arbitrary. But what could those reasons be, other than that those commands are the moral law? But if God does issue those commands because they are the moral law, it looks like explanation (1) above is correct, and divine command theory is false.

Suppose that one responds to Socrates' challenge by saying that God does not command us to follow certain rules because those rules are the moral law — rather, God commands that we follow these rules, but without having any reason for doing so.

This escapes Socrates' challenge, but seems to lead to another problem, which was pressed by the 17th century English philosopher Ralph Cudworth.

According to Cudworth, the following is a consequence of divine command theory:

> nothing can be imagined ſo groſsly wicked, or ſo fouly unjuſt or diſhoneſt, but if it were ſuppoſed to be commanded by this Omnipotent Deity, muſt needs upon that Hypotheſis forthwith become Holy, Juſt and Righteous.

Cudworth is saying that, if divine command theory is true, then, if God had commanded us to murder, cheat, and steal, then murdering, cheating, and stealing would be morally permissible. But surely even if God had commanded us to do these things, they would not be morally permissible!

One might express Cudworth's argument against divine command theory as the following reductio argument:

1. Possibly, God commands that we murder, cheat and steal.
2. Necessarily, if God commands that we do x, then we are morally required to do x.

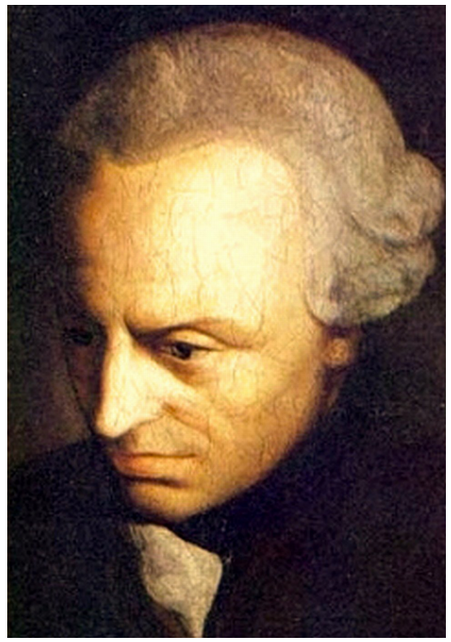
———————————————————

C. Possibly, we are morally required to murder, cheat, and steal.

It does really seem like the conclusion is false, and hence that one of the premises must be false as well.

It is natural to respond to this reductio by rejecting premise (1). But can one reasonably do that if one says, in response to Socrates, that God commands that we follow certain rules for no reason at all?

Let’s turn to our second answer to our question of what makes certain rules part of the moral law: the answer given by the German philosopher Immanuel Kant.



According to Kant, the consequentialist gets things exactly backwards:

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.

What makes a will good is its conformity with a certain rule, which Kant called the categorical imperative.

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 categorical imperative, 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 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.

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.

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.

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 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.

Here is one example of a case in which a consequentialist and a Kantian will say different things. Suppose 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 to so clear.

Here is one such example:

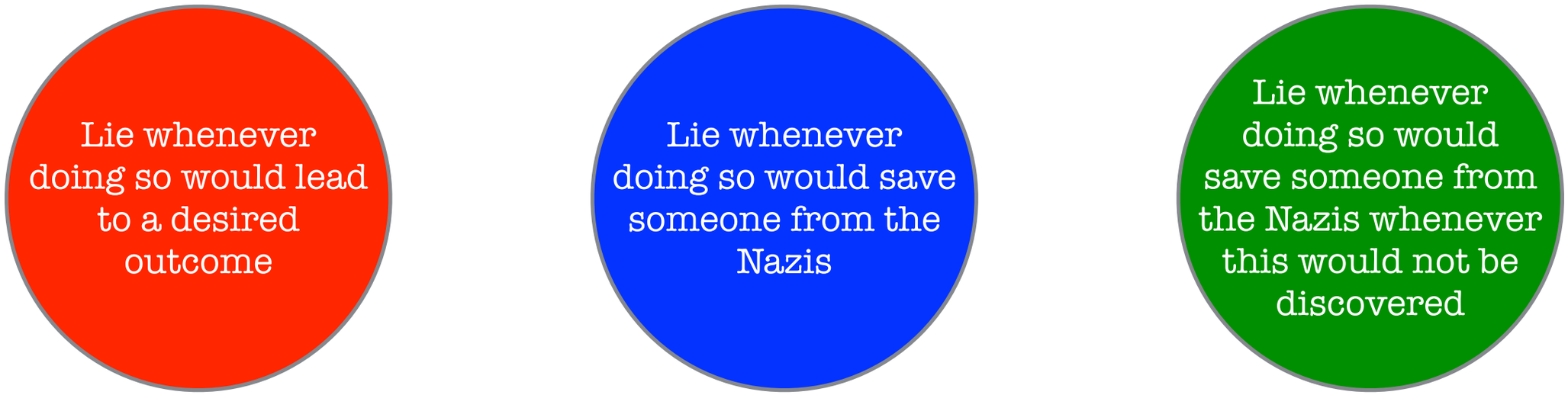
> You're living in Nazi Germany, and hiding a Jewish family in your basement. The authorities come to the door, and ask you whether you are hiding a Jewish family in your house. You know that they will believe you if you tell them that you are not; it is just a random check. What should you do?

What does the Kantian say about this sort of case?

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.

This is one important source of objections to Kant's approach to ethics. In many cases, if the consequences of obeying the categorical imperative are bad enough, many of us feel inclined to say that one ought to break the categorical imperative, in order to avoid the bad consequences.

Let's turn to a second objection to Kant's approach to ethics, which focuses on problems with identifying the maxim out of which someone acts. In the example involving the Nazis above, one might suggest any of the following as the relevant maxim:



The first two maxims seem to fail the formula of universal law. But how about the third?

This difference should be a bit worrying for the Kantian; it is not, after all, easy to see how one could tell which of the second and third is one's real reason for acting. And it is also odd that acting from maxim the second should be morally forbidden, but not acting from the third.

These worries about identifying the relevant maxim are connected with a third 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?

So far we've been considering objections to Kant's use of the formula of universal law. 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. 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?

> *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 mere 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?

One might also worry about its generality — what does it say about our obligations to animals, or the environment? Nothing, it appears.

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 proponent of rule-based ethics, 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 rule-based position.