

# Week 3 Notes

## Philo 101 Online | Hunter College | Fall 2017

Daniel W. Harris

September 9, 2017

### 1 Deontological Ethics

Last week, we focused on one influential family of ethical theories, consequentialism. This week, we will focus on the most influential family of alternatives, which are sometimes grouped together with the term ‘deontological ethics’ or ‘deontology’.

Unlike consequentialist theories, deontological theories don’t judge actions on the basis of their consequences—at least not solely. Deontological theories share the idea that there are some ethical rules or principles so important that we shouldn’t violate them, even if doing so would have the best available consequences. You’ve likely encountered this idea before, in the form of a slogan: “the ends don’t always justify the means”.

It may be easier to understand this using an example. Here’s a moral rule that seems quite bulletproof: *don’t murder innocent people*. This is a good candidate for a moral rule that shouldn’t be violated under any circumstances. So, suppose that we adopt a deontological theory according to which this is one of the inviolable moral rules. But it’s not hard to imagine a situation in which a consequentialist might say that we should violate this rule. Imagine that a super villain has wired a school with explosives, and has threatened to blow it up, killing all of the 25 children who are trapped inside. The villain has demanded that the authorities choose an innocent person at random and kill them as a kind of human sacrifice. If they do it, the villain says, he will let the schoolchildren go. Suppose, further, that the authorities have studied the past behavior of the villain and found that he always keeps his word; there is every reason to think that he would do so again this time. So, what should be done? Murder one innocent person at random in order to save 25 children, or let

them die in order to avoid murdering an innocent? Most consequentialist theories would say that we should sacrifice the life. After all: whatever your way of ranking outcomes, 25 dead children is presumably worse than one dead innocent person. And there is something intuitive about this: although it might feel bad to be the one to pull the trigger, perhaps it would be justified. But any deontological theory that absolutely forbids murdering innocents would say that we should let the children die. The ends don't justify the means. Again, there is something intuitive about this: it is tempting to think that the innocent person who would be chosen has an inviolable right not to be murdered, even if doing so would benefit others. And, indeed, our government would not (publicly) kill an innocent person in order to save lives in this way—at least not unless the person volunteered in some way.

Imaginary (and sometimes real) scenarios like this demonstrate that both consequentialism and deontology both match up with our ordinary moral reasoning some of the time, and seem to pull us in different directions. This is why consequentialism and deontology have been the two most influential families of moral theory: each captures something deep and important about what it takes to live a good life—the pursuit of happiness (for oneself and for others), on one hand, and the inalienable right not to be treated badly by others, on the other.

## 2 An Old-School Deontological Theory

Like consequentialism, deontology is not any single ethical theory, but a whole family of them. To get a particular consequentialist theory, you need a way of ranking the outcomes of possible actions. To get a particular deontological theory, you need some way of choosing one or more moral rules that specify what we ought to do. So, if we want to be deontologists, how should we choose what the rules are?

One option is to simply make a list. Okay, but how should we decide what goes on it? An option that has appealed to have a supernatural being solve this problem for them. You're probably already familiar with one such list, which features in several major religious traditions: the ten commandments. This is a list of central rules that Jews, Christians, and Muslims are never supposed to violate. (The commandments come in different translations and orders, and different books have slightly different lists. I'm not going to choose between the options here. In fact, most of what I'm saying about the ten commandments is disputed and controversial.)

One problem with any such list of rules is two rules on the list might sometimes come into conflict. So, for example, here are two of the commandments:

- Honor your father and your mother.
- Don't kill people.

It's not totally clear what honoring your parents amounts to, but on some influential interpretations, it includes *obeying* them—doing what they tell you to do. So, suppose this is right, and imagine the following scenario. David does his best to follow the commandments, including the commandment to honor his parents, which he interprets as meaning that he must obey them. But one day his parents come to him, give him a gun, and demand that he use it to kill someone whom they have decided is their enemy. What should he do? The answer might seem obvious, but the important thing to see here is that the commandments, on their own, can't decide this issue, because they conflict about it. One of them tells David to obey his parents, and the other one tells David not to do the thing that his parents have demanded that he do.

The fact that the commandments may sometimes come into conflict is not a decisive objection to them. Religious authorities would no doubt have sophisticated ways of solving problems like this. The only point here is that the commandments, on their own, don't constitute a full ethical theory. We also need some way of deciding how to resolve conflicts between them, perhaps by ranking them in order of importance or further interpreting them in some way.

But this illustrates a broader problem about deontological theories, which is that it's quite hard to come up with a list of rules that don't sometimes conflict, and so we also need some way of resolving conflicts. More generally, it would be good to not just have a list of rules, but also some principled way of deciding which rules should be on the list, and which are the most important.

### 3 W. D. Ross's Deontology

Another influential ethical theory that has been formulated as a list of rules is due to the philosopher, W. D. Ross (1877–1971).<sup>1</sup> He came up with the following list of rules, which he called our *prima facie duties*.

- (1) FIDELITY: Keep your promises, and fulfill the agreements into which you have entered.

---

<sup>1</sup>Ross formulated his theory in his book, *The Right and the Good* (1930).

- (2) REPARATION: Make up for the wrongful acts you have previously done to others.
- (3) GRATITUDE: Repay others for past favors done for oneself.
- (4) JUSTICE: Do what is necessary to prevent or correct mismatches between people's pleasure or happiness and what they deserve.
- (5) BENEFICENCE: Do what you can to increase the virtue, intelligence, and happiness of others.
- (6) SELF-IMPROVEMENT: Do what you can to improve your own virtue and intelligence.
- (7) NONMALEFICENCE: Do not harm others.

Ross didn't claim that his list was complete. Other rules might need to be added. And he was open to the possibility that some are consistently more important than others. Plausibly, it is almost always more important not to harm others than to express gratitude, for example. How did he come up with this list? He thought that these rules could be discovered by pure philosophical reflection, by reflecting on the nature of the morally loaded concepts that we use. He argued, for example, that anyone who even understands the concept of a promise should be able to recognize that we have a duty of fidelity to keep our promises.

It's also worth noticing that some of these rules—beneficence, self-improvement, and nonmaleficence—are consequentialist in spirit, since they judge actions by their outcomes. If these were Ross's only rules, then his theory would be a version of consequentialism. But Ross is not a consequentialist, specifically because some of his rules aren't about consequences, and these rules can sometimes outweigh the ones that are. So, for example, Ross thought that the right thing to do is sometimes to keep a promise you have made, even if doing so will ultimately result in more harm than good. (Put another way: when it comes to breaking promises, the ends don't always justify the means.)

Clearly, Ross's rules will sometimes conflict with one another. It's not hard to think of a situation in which you could benefit others (or avoid harming harming them) by breaking a promise, for example. Ross recognized this, and did not claim that these rules are without exceptions. Instead, he thought that, when deciding what to do, we should ask how each of these rules applies to the choice and, in the

case of conflict, weigh them against one another. What is their order of importance? Ross argued that there could be no completely general answer to this question. Consider the conflict between *beneficence*, which instructs us to try to make people happy, and *justice*, which instructs us to correct mismatches between people's happiness and what they deserve. Sometimes, justice trumps beneficence: even if it would make a thief happy to keep what they've stolen, we probably shouldn't let them keep it. But sometimes, beneficence trumps justice: if a child steals one of my french fries, there's no sense in which they "deserve" the happiness they'll get from it, but it would be wrong of me to try to correct this mismatch by taking it out of their mouth.

So, many philosophers have criticized Ross's theory on the ground that it ultimately leaves too much up to our case-by-case moral judgment. It is too vague. And again, this vagueness arises specifically because its rules are hard to interpret and weigh against one another in the case of conflict. This is traditionally the biggest problem for deontological theories.

## 4 Kant's Ethics

The most influential deontological ethical theory, and the one on which the other readings focus, is due to Immanuel Kant (1724–1804). One reason for this influence (though not the only reason) is that Kant's theory attempts to do away with the problem of conflicting rules head-on. The way it does this is by formulating a single, all-important rule—a super-rule that in turn determines which other rules we should follow. Kant's name for this rule is 'the categorical imperative'. And, as Kant makes clear near the beginning of the assigned reading, he is no consequentialist:

What makes a good will good? It isn't what it brings about, its usefulness in achieving some intended end. Rather, good will is good because of how it wills—i.e. it is good in itself. Taken just in itself it is to be valued incomparably more highly than anything that could be brought about by it in the satisfaction of some preference—or, if you like, the sum total of all preferences!

Kant's categorical imperative is a non-consequentialist, universal moral rule. He thinks of it as the instruction that you have to follow in order to count as having a good will.

An imperative is a direction about what to do. Kant calls his rule the *categorical* imperative to distinguish it from *hypothetical* imperatives. These are directions that tell you what to do given that some background condition is met. We can make these background conditions explicit in English using the word ‘if’. Here’s an example: ‘if you want to get a good grade, go to class.’ This imperative doesn’t tell you anything about what to do if you don’t want a good grade. It gives you a direction only relative to the background condition of wanting a good grade. Another way to think about it is that a hypothetical imperative directs you what to do only on the assumption that you want to achieve some further goal. Most imperatives are hypothetical in this way, even if we don’t make the condition explicit. If I tell you ‘study for the next exam,’ for example, we both understand that my direction is irrelevant to you if you don’t care about doing well in the course.

By contrast, a categorical imperative is unconditional. It is an imperative that you must follow, not because doing so would allow you to achieve some further goal that you may or may not have, but because it is the right thing to do in and of itself. This is why Kant thinks that his categorical imperative would be a good foundation for an ethical theory.

What, then, is Kant’s categorical imperative? Kant actually gives several formulations of it. He thinks that they were equivalent—that they would categorize our actions in the same way. However, it is not easy to see how this equivalence works, in part because some of the formulations are themselves difficult to interpret. Here I will mention two of them—both highly influential, the first hard to interpret, and the second a bit more straightforward.

Here’s the first formulation:

CATEGORICAL IMPERATIVE, FIRST FORMULATION

Act only on that maxim through which you can at the same time will that it should become a universal law.

What does it mean to act on a maxim? Well, a maxim is itself a kind of rule or policy that one follows when acting. One of the maxims I try to follow as a professor is to try as hard as I can to make students more interested, rather than less, for example. So, this statement of the categorical imperative is itself a way of choosing maxims—a meta-rule for choosing rules by which to live. It says that we should only live by rules that are rationally universalizable—i.e., that it would make rational sense to want everyone to always follow.

Just what it means for a maxim to be rationally universalizable is a tough question. Many philosophers have thought that it is a very deep idea, though in need of some interpretation, and have spent a lot of energy trying to figure it out. (Remember: this is the hard formulation.) But it's perhaps easiest to see what Kant means by thinking about a maxim that is not rationally universalizable. Here's an example: 'lie as much as possible'. What would happen if everyone tried to act according to this maxim all the time? Well, pretty quickly, we would all stop trusting one another completely, at which point it would become pointless to bother lying. In other words: this is a maxim that, if everyone followed it, would become pointless and impossible to continue following. Clearly, we can't rationally will for that to become a universal law.

This seems intuitive enough. But there are all kinds of hard problems about how to interpret the first version of the categorical imperative. Most importantly, there is the problem that any action can be described as acting on various different maxims. Imagine someone who lies in order to help a friend who is in trouble. This action could be described as one of acting on the maxim, 'tell lies whenever it suits you to do so', or as one acting on the maxim, 'help your friends who are in trouble'. Which one of these maxims is the relevant one here? This is important, because the categorical imperative might tell us that one of them is rationally universalizable and the other is not. So it's often hard to see what the practical upshot of this version of the categorical imperative is supposed to be. Although philosophers have put a lot of thought into how to solve this problem with Kant's first formulation, and some of them have thought that they've been successful, we won't get into their attempts in this course.

Onto the second formulation, which Kant also labels 'the practical imperative', then.

#### CATEGORICAL IMPERATIVE, SECOND FORMULATION

Act in such a way as to treat humanity, whether in your own person or in that of anyone else, always as an end and never merely as a means.

It's much easier to see how we're supposed to follow this rule. It tells us not to treat other people as tools, to be used in whatever way will help us to achieve our goals, even if those goals are otherwise admirable ones. Instead, we should treat others humans as rational, autonomous agents with dignity and a right to make decisions for themselves and to choose their own way to live. It's easy to see why murder, coercion, and dishonesty are all immoral, according to this rule, even if they are

done with good consequences in mind. Each of these kinds of action disrespects the victim and takes away their autonomy. Instead of allowing them to make an informed decision about how to live, these actions treat their victims as things to be manipulated for one's own purposes.

It's should not be hard to think of actions about which utilitarians would disagree with Kant about what to do. According to utilitarians, human dignity and autonomy has no special value in itself, except as a means of increasing happiness. If we can increase the amount of happiness in the world by lying, cheating, coercing, or even killing someone without their consent, utilitarianism says not just that we *can* do so, but that we *must*. Because these actions would involve treating humans as mere means, and not as ends in themselves, Kant would disagree. No ends justify treating humans as mere means, according to Kant.

As Michael Sandel points out in one of this week's required videos, Kant is famously difficult to understand. This is why I've also assigned Sandel's video and Onora O'Neill's essay to compliment the excerpt from Kant's own work. Sandel and O'Neill are among the clearest and most influential contemporary interpreters of Kant's moral philosophy, and they bring out quite different aspects of his ethical theory, which is complex and multidimensional, but worth exploring.

## References

Ross, W. D. (1930). *The Right and the Good*. Oxford University Press, Oxford, UK.