Immanuel Kant and Deontology

LUCAS THORPE

Introduction

This chapter has two main sections. In the first section I briefly sketch Immanuel Kant’s moral theory as laid out in his *Groundwork of the Metaphysics of Morals* (1785). I explain Kant’s claim that morality must be grounded on what he calls a categorical imperative and examine his three formulations of this categorical imperative. In the second section I explain the distinction between “deontological” and “teleological” ethical theories. Kantian ethics is often presented as the paradigm example of a deontological ethical theory, but I question whether Kant’s ethics should be understood as purely deontological.

Kant’s Ethics

Many people think that what is really important in life is to be, say, clever or brave or happy. But Kant thinks that none of these things is good if one lacks a good will: “It is impossible to think of anything at all in the world, or indeed even beyond it, that could be considered good without limitation except a good will . . . a
good will seems to constitute the indispensable condition even of worthiness to be happy” (Groundwork 4:393). For Kant, the most important question in moral philosophy is therefore not “What is it to be happy?” but the question “What is it to have a good will?” And a “good will” is understood not in terms of what it achieves or what it aims at, but the fact that it is moved by—or governed by—reason.²

But Kant thinks that human beings are imperfectly rational beings, and as such our reason is always potentially in conflict with our inclinations. For morally imperfect beings like ourselves, morality is therefore always a matter of duty.³ Truly moral behavior involves acting from duty and not merely acting in accordance with duty (4:397). Take the example of promising to meet my friend at a certain place to help them study for an exam. Perhaps I really enjoy spending time with them, even helping them study for an exam. In such a case I am doing what duty demands, but I am not necessarily doing it from duty. But if I were to still come and help, even if I was not in the mood and would much rather be going to a party, then this would be an example of acting from duty and not merely in accordance with duty.

It is important to note that Kant is not saying here that it would be morally better to help one’s friends without enjoying it. Instead, he is merely pointing out that a good person is one who would choose to keep their promise even if they were not so inclined. It is therefore possible, and desirable, to act from duty and from inclination simultaneously. Ideally a good person does the right thing and enjoys doing it. It is just that a good person would still do the right thing even if they did not have the inclination to do so. This is what it is to act from duty. The person who merely acts in accordance with duty would forget about their duty if their inclination changed. They might keep their promises if they enjoy doing what they promised to do but will break their promises if they thought they would not enjoy keeping them.

Kant further argues that the obligation of duty is to be thought of in terms of an imperative or command. Thus he says, “The representation of an objective principle, insofar as it is necessitating for a will, is called a command (of reason), and the formula of the command is called an imperative. All imperatives are expressed by an ought and indicate by this the relation of an objective law
of reason to a will that by its subjective constitution is not necessarily determined by it (a necessitation)” (4:413). The point Kant is trying to make here is quite simple. Morality is something that binds and constrains us; when we recognize the call of duty, we recognize that there is something that we should do. And the fact that I should do something is not a mere subjective preference, but an objective command. Because of its objective law-like character, then, the demand of morality cannot be explained purely in terms of satisfying one’s desires, because our desires are contingent and subjective, whereas laws are necessary and objective.

Although the content of the moral law is necessary and objective, it does not have the same status as a law of nature. Laws of physics tell us what will be the case; laws of morality only tell us what should be the case. And because we are imperfectly rational and have free will, the fact that we recognize that we should do something does not mean that we will do it. Thus, although the moral law is objectively binding on us, the “subjective constitution” of our will is “not necessarily determined” by the law. In other words, we do not always in fact do what we recognize we ought to.

But what does morality command? To answer this question, Kant distinguishes between what he calls hypothetical imperatives and categorical imperatives. A hypothetical imperative is a conditional (if-then) command. For example, a hypothetical imperative may say: If you want to pass your class, do your homework! Or, If you want to be happy, spend time with your friends! Such hypothetical imperatives only command conditionally. If you do not want to pass the class, then you do not need to do your homework. If you do not want to be happy, you do not need to spend time with your friends. Kant believes, however, that moral obligations cannot be conditional in this way: morality commands absolutely, not conditionally. Thus Kant claims that

Since every practical law represents a possible action as good and thus as necessary for a subject practically determinable by reason, all imperatives are formulae for the determination of action that is necessary in accordance with the principle of a will which is good in some way. Now, if the action would be good merely as a means to something else the imperative is hypothetical; if the action
Morality, Kant thinks, must be based on a categorical imperative rather than a hypothetical one.

Kant then argues that merely reflecting on the concept of what a categorical imperative is allows us to grasp what such an imperative must demand. Thus, in one of the most difficult and most famous paragraphs in the *Groundwork* Kant argues:

> When I think of a hypothetical imperative in general I do not know beforehand what it will contain; I do not know this until I am given the condition. But when I think of a categorical imperative I know at once what it contains.

> For, since the imperative contains, beyond the law, only the necessity that the maxim be in conformity with this law, while the law contains no condition to which it would be limited, nothing is left with which the maxim of action is to conform but the universality of a law as such; and this conformity alone is what the imperative properly represents as necessary. There is, therefore, only a single categorical imperative and it is this: *act only in accordance with that maxim through which you can at the same time will that it become a universal law.* (4:420–21)

The argument in this paragraph is quite abstract, but the basic idea is quite simple: As an objective command, a categorical imperative must be a law. However, as categorical, it must be an unconditioned law. It must be a command that cannot command us to do anything in particular; it can only be a command the only content of which is that it is commandlike. In other words, the categorical imperative must be a purely formal and contentless law, as any content over and above its form would make it conditioned. The only thing that a law whose content is merely the form of being lawful could command is that one act lawfully. So the categorical imperative tells us that an action is only moral if the underlying motivation (maxim) is such that it could serve as a universal law.

Immediately after claiming that there can only be a single categorical imperative, Kant proceeds to introduce what he calls
three formulations of this imperative. There has been much debate as to how, or whether, these distinct formulations can be thought of as formulations of a single imperative. I finish this section by explaining each of these formulations, and then in the final section of the paper I say something about the relationship between them.

In the course of explaining the three formulations of the categorical imperative, Kant uses four concrete examples to illustrate each formulation. These are examples of duties that his contemporaries would have accepted as obvious moral duties. The examples are not chosen randomly, for Kant thinks that there are two ways we can divide duties: between duties to oneself and duties to others, and between what he calls perfect and imperfect duties. Perfect duties command us to perform specific actions. For example, Kant thinks that repaying a debt when asked is a perfect duty, because one has a particular duty to perform a specific action—pay back that particular debt to a particular person. Imperfect duties, by contrast, do not demand specific actions. For example, Kant thinks that we have a duty to promote the happiness of others, but because this duty does not tell us specifically whose happiness to promote and precisely when to do so, it is imperfect. Given these distinctions, we can categorize duties into four classes. Kant discusses one example from each class:

1. **Perfect duties toward oneself**: The duty not to commit suicide.
2. **Perfect duties toward others**: The duty not to make a promise with the intention of breaking it. (The duty to not make a “lying promise.”)
3. **Imperfect duties toward oneself**: The duty to develop one’s talents.
4. **Imperfect duties toward others**: The duty to care about the happiness of others. (The duty of beneficence.)

**Formulations of the Categorical Imperative**

I now briefly explain each formulation of the categorical imperative. The first formulation of the categorical imperative is the
formula of universalizability. This formulation states: “act as if the maxim of your action were to become by your will a universal law of nature” (4:421). Simply, Kant thinks that whenever we act, we have some subjective principle of action (what he calls a “maxim”) that lies behind our action. The formula of universalizability asks us to try and imagine a world in which everyone acted on the same subjective principle that we did; that is, he asks us to imagine that the subjective principle of our action was an objective law.

Kant thinks that in some cases it is not even possible to conceive of a world in which our subjective principle was an objective law. This can be called a contradiction in conception. For example, suppose I am tempted to make a lying promise, with the maxim of my action being the following subjective principle: “when I believe myself to be in need of money, I shall borrow money and promise to repay it, even though I know that this will never happen” (4:442). It is possible for me individually to act on such a maxim, in isolation. But if I try and conceive of a world in which everyone acted on such a maxim, I can see that such a world would not be possible. The ability to make a lying promise presupposes a world in which the institution of promise making exists. If everyone made lying “promises” whenever it was convenient, there could not really be any such thing as promise making. So, Kant concludes, a world in which such a maxim was a universal law is inconceivable, and this formula of the categorical imperative therefore suggests that acting on such a maxim would be immoral.

It is important to stress here that Kant is not saying that maxim is immoral because the consequences of everyone following the rule would be bad, but that it is simply impossible to conceive of a world in which the rule were a law that everyone followed.

There are also contradictions in willing, of which beneficence is a good example. Consider a maxim of never willing to help others when they are in need. Imagine a world in which everyone acted in this way. Although such a world is conceivable, Kant thinks that nobody could will such a world. For we are all beings who are sometimes in need of help from others, and willing a world in which no one was motivated to help others would involve willing a world in which we were never helped when we needed it to achieve our own goals. Such willing would be self-defeating. For
when we aim to achieve something (when we “will an end”), we also will the means to that end. But to will a world in which no one ever helps anyone else would be to will a world in which certain means to our ends (help from others) are not possible. In willing such a world together with willing particular ends, we both will the means to our ends and will that such ends are not available. Hence, our willing is contradictory.

The second formulation of the categorical imperative is the formula of humanity, which states: “So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means” (4:428). Kant introduces this formulation by distinguishing between what he calls subjective ends and objective ends. He argues that moral principles, insofar as they involve ends, have to involve objective ends—ends that hold necessarily for all rational beings. Subjective ends, such as satisfying one’s desires, are ends that depend on the contingent nature of particular agents and as such are ends that can only serve as the basis of hypothetical imperatives. Appealing to such subjective ends, then, can “furnish no universal principles, no principles valid and necessary for all rational beings” (4:428). But, Kant asks, “suppose there were something the existence of which in itself has an absolute worth, something which as an end in itself could be a ground of determinate laws; then in it, and in it alone, would lie the ground of a possible categorical imperative” (4:428). Kant argues that rational beings are such ends in themselves. That is, rational beings are persons who demand respect, not mere things to be used for our own purposes.

To illustrate what is involved in treating others as ends and never merely as means, Kant returns to his four examples. For instance, respecting the humanity of others is not compatible with making a lying promise, as the other person “cannot possibly agree to my way of behaving toward him” (4:430). Kant’s discussion of this example suggests that respect for the humanity of others involves only treating them in ways that they could, at least in principle, agree to.

But, Kant thinks, respect for the humanity of others involves more than just treating people in ways that they can possibly consent to: one must also have some concern with promoting the happiness and well-being of others. Kant also thinks we have a
duty to develop our own talents and moral character, and a duty not to commit suicide, for committing suicide is to fail to treat oneself as an end in itself.

The third formulation of the categorical imperative is sometimes referred to as the formula of the realm (or kingdom) of ends and sometimes as the formula of autonomy. Kant’s discussion of the realm of ends and his introduction of the notion of autonomy is influenced by Rousseau’s account of the idea of an ideal republic in his Social Contract. In such a republic, each individual member is a citizen, and the laws are made by and endorsed by all citizens. As such, each individual citizen is both the source of the laws that govern the community and a subject of those laws. In the Groundwork, Kant takes Rousseau’s political ideal of a republic and turns it into his ethical ideal of a realm of ends. Thus Kant explains that “a rational being belongs as a member to the realm of ends when he gives universal laws in it but is also himself subject to these laws” (4:433).

The third formulation of the categorical imperative says that we should act in a way such that we could be a citizen in such an ideal community. This involves only acting on maxims that could potentially be laws in such a community and respecting other human beings as if they were fellow citizens within such a moral community. This is what Kant means when he claims that “morality consists, then, in the reference of all action to the lawgiving by which alone a realm of ends is possible” (4.434). When we evaluate the maxims behind our actions, we must think of ourselves, and all other human beings, as constituting a single community potentially governed by laws that all could will, and we should ask when evaluating our own maxims whether they could serve as laws for such a community.

Like Rousseau, Kant identifies the idea of being a member of a realm of ends with possession of a certain type of freedom, which he names “autonomy” (from the Greek: autos = self; nomos = law). An individual is autonomous if they are subject to a law they have made themselves. So being a member of a realm of ends is the same as being an autonomous individual who possesses moral freedom. This is why this formulation is sometimes called the formula of autonomy.
Is Kantian Ethics Deontological?

Today, especially among Anglo-American ethical theorists, it is common to contrast deontological with teleological ethical theories, and to suggest that all ethical positions can be classified as either deontological or teleological. Thus, for example, William Frankena offers the following influential definition:

Deontological theories deny what teleological theories affirm. They deny that the right, the obligatory, and the morally good are wholly, whether directly or indirectly a function of what is nonmorally good or of what promotes the greatest balance of good over evil for self, one’s society, or the world as a whole . . . For them the principle of maximizing the balance of good over evil, no matter for whom, is either not a moral criterion or standard at all, or, at least, it is not the only basic or ultimate one. (1973, 15)

This way of classifying ethical theories is, however, relatively recent. The word “deontology” seems to have been coined in the early nineteenth century by the utilitarian philosopher Jeremy Bentham, with its roots coming from the Greek deont- (“that which is binding” or “duty”) and -ology (logos, which means “science” or “theory of”). So based on its etymology, it literally means the science of what ought to be done. Through much of the nineteenth century and into the early twentieth century, the word deontology was just used as a posh way of saying “ethical theory.” In the late nineteenth century, however, many philosophers attempted to introduce ways of categorizing moral theories into broad opposing camps. By the 1930s, many philosophers in the English-speaking world came to think that the broadest distinction was between what they named teleological and deontological theories.

Understood in negative terms, the distinction between deontological and teleological moral theories has to do with their understanding of the relationship between the good and the right. Teleological moral theories start out with some understanding of what is good, and they give an account of what is right in terms
of this good. But the deontologist denies this claim, arguing that there are some things that are right or wrong regardless of whether they promote or produce anything good.

Deontology can also be defined positively in terms of picking out those ethical theories that place moral rules at the center of morality. A deontological moral theory is therefore one that says that morality is a matter of obeying moral rules or principles, such as “do not lie!” or “do not kill!” These principles or rules are a source of obligation even if obeying them will not promote or maximize the good.

Kantian ethics is often presented as the paradigm case of a deontological ethical theory in the Western tradition. Kant places the notion of duty or obligation at the center of his ethics and rejects consequentialism. But should Kant be understood as offering a fully deontological ethics? Although Kant is clearly not a consequentialist, it is not obvious that his theory is completely non-teleological. After all, the concept of a good will, humanity, and the realm of ends are central to his ethical theory, and all seem to be, in a sense, ideas of the good.

Kant certainly does think that moral principles and rules are important: a virtuous individual is one who recognizes and abides by the rules of morality. However, as we have seen, there is a lot more to his ethical thought than merely following rules or principles, and there are disagreements among Kant scholars about the degree to which moral rules and principles play a foundational role in his ethical system.

Perhaps the most influential recent interpretation of Kant’s ethics is offered by John Rawls, who offers a strongly deontological reading of Kant’s ethics. He names this interpretation “Kantian constructivism.” A central commitment of the constructivist position is the deontologist’s claim that the “the right is prior to the good.” To understand what this slogan means, we can contrast the constructivist position with that of the utilitarian, who believes, in contrast, that the “good is prior to the right.” Thus Rawls explains that “Utilitarianism starts with a conception of the good given prior to, and independent of, the right (the moral law), and it then works out from that independent conception its conceptions of the right and of moral worth, in that order” (1989, 92). In other words, first the utilitarian works out what goods we should be
aiming at. Once she knows this, she can then discover what we ought to do (what is “right”) by seeing whether the action brings about those goods.

A Kantian constructivist, by contrast, starts with a conception of the right and defines the good in terms of the right: a state of affairs is good if it was (or perhaps could have been) chosen in the right way. This disagreement can be understood as a disagreement about what it is to be reasonable. The utilitarian will define reasonableness in terms of the good, whereas the constructivist will define the good in terms of the reasonable.

For instance, imagine a group of individuals who wish to share a cake. The utilitarian will argue that the procedure we use to determine how to divide up the cake is reasonable if it is intended or likely to produce the best decision; that is, a decision that maximizes total happiness. The constructivist, by contrast, will argue that a decision is good if the procedure used to make the decision was a reasonable one. On this approach, what it is to be reasonable must be defined independently of, and prior to, any conception of the good or the desirable.

Rawls’s constructivist interpretation is based on an account of the relationship between the first and third formulation of the categorical imperative. He interprets the first formulation as providing an account of the right and the third formulation as providing an account of the good.

A constructivist ethics starts by providing an account of the “reasonably willable.” This is precisely what Rawls believes Kant is trying to do when he introduces the first formulation of the categorical imperative in the *Groundwork*. Rawls argues that the first formulation of the categorical imperative should be understood as introducing a procedure to test the reasonableness of maxims, which he calls the CI-Procedure (2000, 181). According to Rawls, the first formulation of the categorical imperative is an attempt to give an account of what it is to be reasonable in purely deontological terms, terms that make no reference to the good or the morally desirable. To be reasonable is to implicitly follow a certain procedure—the CI-Procedure—and the good or morally desirable is to be defined in terms of this procedure.

If the first formulation of the categorical imperative specifies a procedure to test the reasonableness of our maxims, the second
two formulations specify the objects that such a reasonable person
should value (or find good). A reasonable person will value the
humanity of herself and others and will value the idea of being
a member of a realm of ends. The notions of “humanity” and “a
realm of ends” are concepts of the good. Given his reading of Kant
as a moral constructivist, Rawls believes that Kant is committed to
the position that these two ideas must be defined in terms of the
procedure introduced in the first formulation. In other words, he
believes that the second and third formulations of the categorical
imperative are dependent for their content on the first formula-
tion. Rawls and his followers, then, read Kant as committed to
a strongly deontological position in both the negative sense of
rejecting teleology and in the positive sense of putting moral rules
(the CI procedure) at the heart of his ethics.

Critics of such a strongly deontological interpretation tend
to stress the second and third formulations of the categorical
imperative, arguing that these formulations offer some notion of
the good independent of the right. Such approaches stress the idea
that Kant thinks that human beings are ends in themselves and
are deserving of respect regardless of whether anyone actually
chooses to show such respect. According to this more teleological
interpretation, Kantian ethics presupposes the value of humanity
and the value of being a member of citizen in a moral community.

Kant begins the *Groundwork* by examining our everyday eth-
ical commitments. He thinks that it is a part of our commonsense
morality that we should only act on principles that can be univer-
salized. But then he asks: why do we care about universalizability?
And his answer is that we care about universalizability because
we care about other human beings and about standing in the right
sort of relationship with them. According to this more teleologi-
cal interpretation of Kant’s ethics, the structure of the argument
has to do with uncovering what we ultimately value. We do not
construct an idea of the good from a set of (meaningless) moral
rules. Instead, we care about certain moral rules because we rec-
ognize the value of other human beings and our standing in the
right sort of relationship to them. If this interpretation is correct,
then Kantian ethics is ultimately teleological (and not fully deon-
tological) because the value of humanity and the value of being a
member of a realm of ends are conceptions of the good and are foundational for his ethics.

There is, then, disagreement between scholars about the degree to which Kantian ethics should be understood in purely deontological terms, and to what degree notions of the good play a foundational role in his ethical theory. 5

Summary

Kant’s argument in the *Groundwork* that morality must involve a categorical imperative, and his account of the three formulations, has probably been his most influential contribution to ethics in the Western tradition. His appeal to universalization as a criterion to judge the validity of principles of action has become one of the most influential alternatives to utilitarian accounts of morality. And Kant’s appeal to the value of humanity has played a large role in the development of thinking about human rights, with many defenders of human rights attempting to ground such rights in Kantian terms: in the dignity of human beings. In addition, many have found the idea that we should treat others as ends rather than means an attractive moral ideal; for example, feminist critiques of the “objectification” of women can be understood in such Kantian terms. Finally, Kant’s appeal to the value of autonomy has also been extremely influential. However, there has been a regrettable tendency to understand Kant’s notion of autonomy in individualistic terms (as a capacity for self-determination) rather than in social terms (as a capacity to be a citizen in an ideal moral community). 6

Notes

1. Support for work on this paper was provided by Boğaziçi University Research Fund Grant Number 15681. Thanks to Michael Hemmingsen for comments on various drafts of this chapter.

2. What exactly Kant means by reason is a bone of contention among Kant scholars, and much of his practical philosophy can be understood as an attempt to clarify what is involved in being governed by reason. Some understand reason primarily in terms of a capacity of rationality
and rules. Others understand reason to involve a capacity to set ends
and pursue them. I take practical reason to essentially involve a (moral)
capacity to interact with others on the basis of mutual respect. See Thorpe
(2018).

3. Kant often imagines the idea of a perfectly rational being, which
he calls the idea of a holy will. For such a being morality would not be
a matter of duty; a morally perfect individual would just act morally
with no inner conflict.

4. There is debate over why Kant thinks that human beings are ends-
in-themselves. One suggestion is that what makes us ends in ourselves is
our capacity to set ends, for only a being that has value has the capacity
to make decisions about what has value.

5. See Tilev (2021) for an account that tries to combine deontological
and teleological conceptions of autonomy.

6. See Thorpe (2011) and Vatansever (2021) for a further discussion
of this point.

Further Reading

Primary Sources

Translation in *Practical Philosophy (The Cambridge Edition of the Works
of Immanuel Kant).* Edited by M. Gregor. Cambridge: Cambridge
University Press.

———. (1788) 1996. *Critique of Practical Reason.* Translation in *Practical Phi-
losophy (The Cambridge Edition of the Works of Immanuel Kant).* Edited
by M. Gregor. Cambridge: Cambridge University Press.

———. (1797) 1996. *The Metaphysics of Morals.* Translation in *Practical Phi-
losophy (The Cambridge Edition of the Works of Immanuel Kant).* Edited
by M. Gregor. Cambridge: Cambridge University Press.

There are so many editions of the *Groundwork* that it is not
very helpful to give the page number of a particular English
edition. Instead, when scholars want to refer to a particular
page in Kant’s work, we refer to the volume number and
the page number of the German edition of Kant’s complete
works. So a reference to the Groundwork of the form 4:443
Most English editions of the *Groundwork* have these numbers
down the side of the page—so whatever English edition you
are using, you should be able to find the passage referred to.
SECONDARY SOURCES


