

Immanuel Kant



At the foundation of Kant's system is the doctrine of "transcendental idealism," which emphasizes a distinction between what we can experience (the natural, observable world) and what we cannot ("supersensible" objects such as God and the soul). Kant argued that we can only have knowledge of things we can experience. Accordingly, in answer to the question, "What can I know?" Kant replies that we can know the natural, observable world, but we cannot, however, have answers to many of the deepest questions of metaphysics.

Kant's ethics are organized around the notion of a "categorical imperative," which is a universal ethical principle stating that one should always respect the humanity in others, and that one should only act in accordance with rules that could hold for everyone. Kant argued that the moral law is a truth of reason, and hence that all rational creatures are bound by the same moral law. Thus in answer to the question, "What should I do?" Kant replies that we should act rationally, in accordance with a universal moral law.

Kant also argued that his ethical theory requires *belief* in free will, God, and the immortality of the soul. Although we cannot have *knowledge* of these things, reflection on the moral law leads to a justified belief in them, which amounts to a kind rational faith. Thus in answer to the question, "What may I hope?" Kant replies that we may hope that our souls are immortal and that there really is a God who designed the world in accordance with principles of justice.

Kant also argued that his ethical theory requires *belief* in free will, God, and the immortality of the soul. Although we cannot have *knowledge* of these things, reflection on the moral law leads to a justified belief in them, which amounts to a kind rational faith. Thus in answer to the question, "What may I hope?" Kant replies that we may hope that our souls are immortal and that there really is a God who designed the world in accordance with principles of justice.

In addition to these three focal points, Kant also made lasting contributions to nearly all areas of philosophy. His aesthetic theory remains influential among art critics. His theory of knowledge is required reading for many branches of analytic philosophy. The cosmopolitanism behind his political theory colors discourse about globalization and international relations. And some of his scientific contributions are even considered intellectual precursors to several ideas in contemporary cosmology.

5. Moral Theory

Kant's moral theory is organized around the idea that to act morally and to act in accordance with reason are one and the same. In virtue of being a rational agent (that is, in virtue of possessing practical reason, reason which is interested and goal-directed), one is obligated to follow the moral law that practical reason prescribes. To do otherwise is to act irrationally. Because Kant places his emphasis on the duty that comes with being a rational agent who is cognizant of the moral law, Kant's theory is considered a form of deontology (*deon*– comes from the Greek for “duty” or “obligation”).

Like his theoretical philosophy, Kant's practical philosophy is *a priori*, formal, and universal: the moral law is derived non-empirically from the very structure of practical reason itself (its *form*), and since all rational agents share the same practical reason, the moral law binds and obligates everyone equally. So what is this moral law that obligates all rational agents universally and *a priori*? The moral law is determined by what Kant refers to as the Categorical Imperative, which is the general principle that demands that one respect the humanity in oneself and in others, that one not make an exception for oneself when deliberating about how to act, and in general that one only act in accordance with rules that everyone could and should obey.

Although Kant insists that the moral law is equally binding for all rational agents, he also insists that the bindingness of the moral law is *self-imposed*: we *autonomously* prescribe the moral law to ourselves. Because Kant thinks that the kind of autonomy in question here is only possible under the presupposition of a transcendently free basis of moral choice, the constraint that the moral law places on an agent is not only consistent with freedom of the will, it requires it. Hence, one of the most important aspects of Kant's project is to show that we are justified in presupposing that our morally significant choices are grounded in a transcendental freedom (the very sort of freedom that Kant argued we *could not* prove through mere "theoretical" or "speculative" reason; see 2gii above).

This section aims to explain the structure and content of Kant's moral theory (5a-b), and also Kant's claims that belief in freedom, God, and the immortality of the soul are necessary "postulates" of practical reason (5c). (On the relation between Kant's moral theory and his aesthetic theory, see 7c below.)

a. The Good Will and Duty

Kant lays out the case for his moral theory in *Groundwork for the Metaphysics of Morals* (1785), *Critique of Practical Reason* (also known as the "Second Critique"; 1788), and the *Metaphysics of Morals* (1797). His arguments from the *Groundwork* are his most well-known and influential, so the following focuses primarily on them.

Kant begins his argument from the premise that a moral theory must be grounded in an account of what is *unconditionally good*. If something is merely conditionally good, that is, if its goodness depends on something else, then that other thing will either be merely conditionally good as well, in which case its goodness depends on yet another thing, or it will be unconditionally good. All goodness, then, must ultimately be traceable to something that is unconditionally good. There are many things that we typically think of as good but that are not truly unconditionally good. Beneficial resources such as money or power are often good, but since these things can be used for *evil* purposes, their goodness is conditional on the use to which they are put. Strength of character is generally a good thing, but again, if someone uses a strong character to successfully carry out evil plans, then

the strong character is not good. Even happiness, according to Kant, is not unconditionally good. Although all humans universally desire to be happy, if someone is happy but does not deserve their happiness (because, for instance, their happiness results from stealing from the elderly), then it is not good for the person to be happy. Happiness is only good on the condition that the happiness is deserved.

Kant argues that there is only one thing that can be considered unconditionally good: *a good will*. A person has a good will insofar as they form their intentions on the basis of a self-conscious respect for the moral law, that is, for the rules regarding what a rational agent ought to do, one's duty. The value of a good will lies in the principles on the basis of which it forms its intentions; it does *not* lie in the consequences of the actions that the intentions lead to. This is true even if a good will never leads to any desirable consequences at all: "Even if... this will should wholly lack the capacity to carry out its purpose... then, like a jewel, it would still shine by itself, as something that has its full worth in itself" (4:393). This is in line with Kant's emphasis on the *unconditional* goodness of a good will: if a will were evaluated in terms of its consequences, then the goodness of the will would depend on (that is, would be *conditioned* on) those consequences. (In this respect, Kant's deontology is in stark opposition to consequentialist moral theories, which base their moral evaluations on the consequences of actions rather than the intentions behind them.)

b. The Categorical Imperative

If a good will is one that forms its intentions on the basis of correct principles of action, then we want to know what sort of principles these are. A principle that commands an action is called an "imperative." Most imperatives are "hypothetical imperatives," that is, they are commands that hold only if certain conditions are met. For instance: "if you want to be a successful shopkeeper, then cultivate a reputation for honesty." Since hypothetical imperatives are conditioned on desires and the intended consequences of actions, they cannot serve as the principles that determine the intentions and volitions of an unconditionally good will. Instead, we require what Kant calls a "categorical imperative." Where hypothetical imperatives take the form, "if y is desired/intended/sought, do x," categorical imperatives simply take the form, "do x." Since a categorical imperative is stripped of all reference to the consequences of an action, it is thereby stripped of all determinate content, and hence it is purely *formal*. And since it is unconditional, it holds *universally*. Hence a categorical imperative expresses only the very form of a universally binding law: "nothing is left but the conformity of actions as such with universal law" (4:402). To act morally, then, is to form one's intentions on the basis of the very idea of a universal principle of action.

This conception of a categorical imperative leads Kant to his first official formulation of the categorical imperative itself: "act only in accordance with that maxim through which you can at the same time will that it become a universal law" (4:421). A maxim is a general rule that can be used to determine particular courses of actions in particular circumstances. For instance, the maxim "I

shall lie when it will get me out of trouble” can be used to determine the decision to lie about an adulterous liaison. The categorical imperative offers a decision procedure for determining whether a given course of action is in accordance with the moral law. After determining what maxim one would be basing the action in question on, one then asks whether it would be possible, given the power (in an imagined, hypothetical scenario), to choose that everyone act in accordance with that same maxim. If it is possible to will that everyone act according to that maxim, then the action under consideration is morally permissible. If it is not possible to will that everyone act according to that maxim, the action is morally impermissible. Lying to cover up adultery is thus immoral because one cannot will that everyone act according to the maxim, “I shall lie when it will get me out of trouble.” Note that it is not simply that it would be *undesirable* for everyone to act according to that maxim. Rather, it would be *impossible*. Since everyone would know that everyone else was acting according to that maxim, there would never be the presupposition that anyone was telling the truth; the very act of lying, of course, requires such a presupposition on the part of the one being lied to. Hence, the state of affairs where everyone lies to get out of trouble can never arise, so it cannot be willed to be a universal law. It fails the test of the categorical imperative.

The point of Kant’s appeal to the universal law formulation of the categorical imperative is to show that an action is morally permissible only if the maxim on which the action is based could be affirmed as a universal law that everyone obeys without exception. The mark of immorality, then, is that one makes an exception for oneself. That is, one acts in a way that they would not want everyone else to. When someone chooses to lie about an adulterous liaison, one is implicitly thinking, “in general people should tell the truth, but in this case I will be the exception to the rule.”

Kant’s first formulation of the categorical imperative describes it in terms of the very form of universal law itself. This formal account abstracts from any specific *content* that the moral law might have for living, breathing human beings. Kant offers a second formulation to address the material side of the moral law. Since the moral law has to do with actions, and all actions are by definition teleological (that is, goal-directed), a material formulation of the categorical imperative will require an appeal to the “ends” of human activity. Some ends are merely instrumental, that is, they are sought only because they serve as “means” towards further ends. Kant argues that the moral law must be aimed at an end that is not merely instrumental, but is rather an end *in itself*. Only rational agents, according to Kant, are ends in themselves. To act morally is thus to respect rational agents as ends in themselves. Accordingly, the categorical imperative can be reformulated as follows: “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” (4:429). The basic idea here is that it is immoral to treat someone as a *thing* of merely instrumental value; *persons* have an intrinsic (non-instrumental) value, and the moral law demands that we respect this intrinsic value. To return to the example of the previous paragraphs, it would be wrong to lie about an adulterous liaison because by withholding the truth one is manipulating the other person to make things easier for oneself; this sort of

manipulation, however, amounts to treating the other as a thing (as a mere means to the comfort of not getting in trouble), and not as a person deserving of respect and entitled to the truth.

The notion of a universal law provides the *form* of the categorical imperative and rational agents as ends in themselves provide the *matter*. These two sides of the categorical imperative are combined into yet a third formulation, which appeals to the notion of a “kingdom of ends.” A kingdom of ends can be thought of as a sort of perfectly just utopian ideal in which all citizens of this kingdom freely respect the intrinsic worth of the humanity in all others because of an autonomously self-imposed recognition of the bindingness of the universal moral law for all rational agents. The third formulation of the categorical imperative is simply the idea that one should act in whatever way a member of this perfectly just society would act: “act in accordance with the maxims of a member giving universal laws for a merely possible kingdom of ends” (4:439). The idea of a kingdom of ends is an *ideal* (hence the “merely possible”). Although humanity may never be able to achieve such a perfect state of utopian coexistence, we can at least strive to approximate this state to an ever greater degree.

c. Postulates of Practical Reason

In *Critique of Pure Reason*, Kant had argued that although we can acknowledge the bare logical possibility that humans possess free will, that there is an immortal soul, and that there is a God, he also argued that we can never have positive knowledge of these things (see 2g above). In his ethical writings, however, Kant complicates this story. He argues that despite the theoretical impossibility of knowledge of these objects, *belief* in them is nevertheless a precondition for moral action (and for practical cognition generally). Accordingly, freedom, immortality, and God are “postulates of practical reason.” (The following discussion draws primarily on *Critique of Practical Reason*.)

We will start with freedom. Kant argues that morality and the obligation that comes with it are only possible if humans have free will. This is because the universal laws prescribed by the categorical imperative presuppose *autonomy* (*autos* = self; *nomos* = law). To be autonomous is to be the free ground of one’s own principles, or “laws” of action. Kant argues that if we presuppose that humans are rational and have free will, then his entire moral theory follows directly. The problem, however, lies in justifying the belief that we are free. Kant had argued in the Second Analogy of Experience that every event in the natural world has a “determining ground,” that is, a cause, and so all human actions, as natural events, themselves have deterministic causes (see 2f above). The only room for freedom of the will would lie in the realm of things in themselves, which contains the noumenal correlate of my phenomenal self. Since things in themselves are unknowable, I can never look to them to get evidence that I possess transcendental freedom. Kant gives at least two arguments to justify belief in freedom as a precondition of his moral theory. (There is a great deal of controversy among commentators regarding the exact form of his arguments, as well as their success. It will not

be possible to adjudicate those disputes in any detail here. See Section 10 (References and Further Readings) for references to some of these commentaries.)

In the *Groundwork*, Kant suggests that the presupposition that we are free follows as a consequence of the fact that we have practical reason and that we think of ourselves as practical agents. Any time I face a choice that requires deliberation, I must consider the options before me as really open. If I thought of my course of action as already determined ahead of time, then there would not really be any choice to make. Furthermore, in taking my deliberation to be real, I also think of the possible outcomes of my actions as *caused by me*. The notion of a causality that originates in the self is the notion of a free will. So the very fact that I do deliberate about what actions I will take means that I am presupposing that my choice is real and hence that I am free. As Kant puts it, all practical agents act “under the idea of freedom” (4:448). It is not obvious that this argument is strong enough for Kant’s purposes. The position seems to be that I must act *as though* I am free, but acting as though I am free in no way entails that I really *am* free. At best, it seems that since I act *as though* I am free, I thereby must act *as though* morality really does obligate me. This does not establish that the moral law *really does* obligate me.

In the *Second Critique*, Kant offers a different argument for the reality of freedom. He argues that it is a brute “fact of reason” (5:31) that the categorical imperative (and so morality generally) obligates us as rational agents. In other words, all rational agents are at least implicitly conscious of the bindingness of the moral law on us. Since morality requires freedom, it follows that if morality is real, then freedom must be real too. Thus this “fact of reason” allows for an inference to the reality of freedom. Although the conclusion of this argument is stronger than the earlier argument, its premise is more controversial. For instance, it is far from obvious that all rational agents are conscious of the moral law. If they were, how come no one discovered this exact moral law before 1785 when Kant wrote the *Groundwork*? Equally problematic, it is not clear why this “fact of reason” should count as *knowledge* of the bindingness of the moral law. It may just be that we cannot help but believe that the moral law obligates us, in which case we once again end up merely acting *as though* we are free and *as though* the moral law is real.

Again, there is much debate in the literature about the structure and success of Kant’s arguments. It is clear, however, that the success of Kant’s moral project stands or falls with his arguments for freedom of the will, and that the overall strength of this theory is determined to a high degree by the epistemic status of our belief in our own freedom.

Kant’s arguments for immortality and God as postulates of practical reason presuppose that the reality of the moral law and the freedom of the will have been established, and they also depend on the principle that “‘ought’ implies ‘can’”: one cannot be obligated to do something unless the thing in question is doable. For instance, there is no sense in which I am obligated to single-handedly solve global poverty, because it is not within my power to do so. According to Kant, the ultimate

aim of a rational moral agent should be to become *perfectly* moral. We are obligated to strive to become ever more moral. Given the “ought implies can” principle, if we *ought* to work towards moral perfection, then moral perfection must be *possible* and we *can* become perfect. However, Kant holds that moral perfection is something that finite rational agents such as humans can only progress towards, but not actually attain in any finite amount of time, and certainly not within any one human lifetime. Thus the moral law demands an “endless progress” towards “complete conformity of the will with the moral law” (5:122). This endless progress towards perfection can only be demanded of us if our own existence is endless. In short, one’s belief that one should strive towards moral perfection presupposes the belief in the immortality of the soul.

In addition to the “ought implies can” principle, Kant’s argument about belief in God also involves an elaboration of the notion of the “highest good” at which all moral action aims (at least indirectly). According to Kant, the highest good, that is, the most perfect possible state for a community of rational agents, is not only one in which all agents act in complete conformity with the moral law. It is also a state in which these agents are *happy*. Kant had argued that although everyone naturally desires to be happy, happiness is only good when one deserves to be happy. In the ideal scenario of a morally perfect community of rational agents, *everyone* deserves to be happy. Since a deserved happiness is a good thing, the highest good will involve a situation in which everyone acts in complete conformity with the moral law and everyone is completely happy because they deserve to be. Now since we are obligated to work towards this highest good, this complete, universal, morally justified happiness must be possible (again, because “ought” implies “can”). This is where a puzzle arises. Although happiness is connected to morality at the *conceptual* level when one deserves happiness, there is no *natural* connection between morality and happiness. Our happiness depends on the natural world (for example, whether we are healthy, whether natural disasters affect us), and the natural world operates according to laws that are completely separate from the laws of morality. Accordingly, acting morally is in general no guarantee that nature will make it possible for one to be happy. If anything, behaving morally will often decrease one’s happiness (for doing the right thing often involves doing the uncomfortable, difficult thing). And we all have plenty of empirical evidence from the world we live in that often bad things happen to good people and good things happen to bad people. Thus if the highest good (in which happiness is proportioned to virtue) is possible, then somehow there must be a way for the laws of nature to *eventually* lead to a situation in which happiness is proportioned to virtue. (Note that since at this point in the argument, Kant takes himself to have established immortality as a postulate of practical reason, this “eventually” may very well be far in the future). Since the laws of nature and the laws of morality are completely separate on their own, the only way that the two could come together such that happiness ends up proportioned to virtue would be if the ultimate cause and ground of nature set up the world in such a way that the laws of nature would eventually lead to the perfect state in question. Therefore, the possibility of the highest good requires the presupposition that the cause of the world is intelligent and powerful enough to set nature up in the right way, and also that it wills in accordance with justice that eventually the laws of nature will indeed lead to a state in which the

happiness of rational agents is proportioned to their virtue. This intelligent, powerful, and just cause of the world is what traditionally goes by the name of “God.” Hence God is a postulate of practical reason.