

Kant's Moral Philosophy

First published Mon Feb 23, 2004; substantive revision Thu Jul 7, 2016

Immanuel Kant (1724–1804) argued that the supreme principle of morality is a standard of rationality that he dubbed the “Categorical Imperative” (CI). Kant characterized the CI as an objective, rationally necessary and unconditional principle that we must always follow despite any natural desires or inclinations we may have to the contrary. All specific moral requirements, according to Kant, are justified by this principle, which means that all immoral actions are irrational because they violate the CI. Other philosophers, such as Hobbes, Locke and Aquinas, had also argued that moral requirements are based on standards of rationality. However, these standards were either instrumental principles of rationality for satisfying one’s desires, as in Hobbes, or external rational principles that are discoverable by reason, as in Locke and Aquinas. Kant agreed with many of his predecessors that an analysis of practical reason reveals the requirement that rational agents must conform to instrumental principles. Yet he also argued that conformity to the CI (a non-instrumental principle), and hence to moral requirements themselves, can nevertheless be shown to be essential to rational agency. This argument was based on his striking doctrine that a rational will must be regarded as autonomous, or free, in the sense of being the author of the law that binds it. The fundamental principle of morality — the CI — is none other than the law of an autonomous will. Thus, at the heart of Kant’s moral philosophy is a conception of reason whose reach in practical affairs goes well beyond that of a Humean ‘slave’ to the passions. Moreover, it is the presence of this self-governing reason in each person that Kant thought offered decisive grounds for viewing each as possessed of equal worth and deserving of equal respect.

Kant’s most influential positions in moral philosophy are found in *The Groundwork of the Metaphysics of Morals* (hereafter, “*Groundwork*”) but he developed, enriched, and in some cases modified those views in later works such as *The Critique of Practical Reason*, *The Metaphysics of Morals*, *Anthropology from a Pragmatic Point of View*, *Religion within the Boundaries of Mere Reason* as well as his essays on history and related topics. Kant’s *Lectures on Ethics*, which were lecture notes taken by three of his students on the courses he gave in moral philosophy, also include relevant material for understanding his views. We will mainly focus on the foundational doctrines of the *Groundwork*, even though in recent years some scholars have become dissatisfied with this standard approach to Kant’s views and have turned their attention to the later works. We find the standard approach most illuminating, though we will highlight important positions from the later works where needed.

- [1. Aims and Methods of Moral Philosophy](#)
- [2. Good Will, Moral Worth and Duty](#)
- [3. Duty and Respect for Moral Law](#)
- [4. Categorical and Hypothetical Imperatives](#)
- [5. The Formula of the Universal Law of Nature](#)
- [6. The Humanity Formula](#)
- [7. The Autonomy Formula](#)
- [8. The Kingdom of Ends Formula](#)
- [9. The Unity of the Formulas](#)
- [10. Autonomy](#)
- [11. Virtue and Vice](#)
- [12. Normative Ethical Theory](#)
- [13. Teleology or Deontology?](#)
- [14. Metaethics](#)
- [Bibliography](#)
- [Academic Tools](#)
- [Other Internet Resources](#)
- [Related Entries](#)

1. Aims and Methods of Moral Philosophy

The most basic aim of moral philosophy, and so also of the *Groundwork*, is, in Kant's view, to "seek out" the foundational principle of a "metaphysics of morals," which Kant understands as a system of *a priori* moral principles that apply the CI to human persons in all times and cultures. Kant pursues this project through the first two chapters of the *Groundwork*. He proceeds by analyzing and elucidating commonsense ideas about morality, including the ideas of a "good will" and "duty". The point of this first project is to come up with a precise statement of the principle or principles on which all of our ordinary moral judgments are based. The judgments in question are supposed to be those that any normal, sane, adult human being would accept on due rational reflection. Nowadays, however, many would regard Kant as being overly optimistic about the depth and extent of moral agreement. But perhaps he is best thought of as drawing on a moral viewpoint that is very widely shared and which contains some general judgments that are very deeply held. In any case, he does not appear to take himself to be primarily addressing a genuine moral skeptic such as those who often populate the works of moral philosophers, that is, someone who doubts that she has any reason to act morally and whose moral behavior hinges on a rational proof that philosophers might try to give. For instance, when, in the third and final chapter of the *Groundwork*, Kant takes up his second fundamental aim, to "establish" this foundational moral principle as a demand of each person's own rational will, his conclusion apparently falls short of answering those who want a proof that we really are bound by moral requirements. He rests this second project on the position that we — or at least creatures with rational wills — possess autonomy. The argument of this second project does often appear to try to reach out to a metaphysical fact about our wills. This has led some readers to the conclusion that he is, after all, trying to justify moral requirements by appealing to a fact — our autonomy — that even a moral skeptic would have to recognize.

Kant's analysis of the common moral concepts of "duty" and "good will" led him to believe that we are free and autonomous as long as morality, itself, is not an illusion. Yet in the *Critique of Pure Reason*, Kant also tried to show that every event has a cause. Kant recognized that there seems to be a deep tension between these two claims: If causal determinism is true then, it seems, we cannot have the kind of freedom that morality presupposes, which is "a kind of causality" that "can be active, independently of alien causes *determining it*" (G 4:446).

Kant thought that the only way to resolve this apparent conflict is to distinguish between *phenomena*, which is what we know through experience, and *noumena*, which we can consistently think but not know through experience. Our knowledge and understanding of the empirical world, Kant argued, can only arise within the limits of our perceptual and cognitive powers. We should not assume, however, that we know all that may be true about "things in themselves," although we lack the "intellectual intuition" that would be needed to learn about such things.

These distinctions, according to Kant, allow us to resolve the "antinomy" about free will by interpreting the "thesis" that free will is possible as about noumena and the "antithesis" that every event has a cause as about phenomena. Morality thus presupposes that agents, in an incomprehensible "intelligible world," are able to make things happen by their own free choices in a "sensible world" in which causal determinism is true.

Many of Kant's commentators, who are skeptical about these apparently exorbitant metaphysical claims, have attempted to make sense of his discussions of the intelligible and sensible worlds in less metaphysically demanding ways. On one interpretation (Hudson 1994), one and the same act can be described in wholly physical terms (as an appearance) and also in irreducibly mental terms (as a thing in itself). On this compatibilist picture, all acts are causally determined, but a free act is one that can be described as determined by irreducibly mental causes, and in particular by the causality of reason. A second interpretation holds that the intelligible and sensible worlds are used as metaphors for two ways of conceiving of one and the same world (Korsgaard 1996; Allison 1990; Hill 1989a, 1989b). When we are engaging in scientific or empirical investigations, we often take up a perspective in which we think of things as subject to natural causation, but when we deliberate, act, reason and judge, we often take up a different perspective, in which we think of ourselves and others as agents who are not determined by natural causes. When we take up this latter, practical, standpoint, we need not believe that we or others really are free, in any deep metaphysical sense; we need only operate "under the idea of freedom" (G 4:448). Controversy persists, however, about whether Kant's conception of freedom requires a "two worlds" or "two perspectives" account of the sensible and intelligible worlds (Guyer 1987, 2009; Langton 2001; Kohl 2016; Wood 1984; Hogan 2009).

Although the two most basic aims Kant saw for moral philosophy are to seek out and establish the supreme principle of morality, they are not, in Kant's view, its only aims. Moral philosophy, for Kant, is most fundamentally addressed to the first-person, deliberative question, "What ought I to do?", and an answer to that question requires much more than delivering or justifying the fundamental principle of morality. We also need some account, based on this principle, of the nature and extent of the specific moral duties that apply to us. To this end, Kant employs his findings from the *Groundwork* in *The Metaphysics of Morals*, and offers a categorization of our basic moral duties to ourselves and others. In addition, Kant thought that moral philosophy should characterize and explain the demands that morality makes on human psychology and forms of human social interaction. These topics, among others, are addressed in central chapters of the second *Critique*, the *Religion* and again in the *Metaphysics of Morals*, and are perhaps given a sustained treatment in *Anthropology from a Pragmatic Point of View*. Further, a satisfying answer to the question of what one ought to do would have to take into account any political and religious requirements there are. Each of these requirements turn out to be, indirectly at least, also moral obligations for Kant, and are discussed in the *Metaphysics of Morals* and in *Religion*. Finally, moral philosophy should say something about the ultimate end of human endeavor, the Highest Good, and its relationship to the moral life. In the *Critique of Practical Reason*, Kant argued that this Highest Good for humanity is complete moral virtue together with complete happiness, the former being the condition of our deserving the latter. Unfortunately, Kant noted, virtue does not insure wellbeing and may even conflict with it. Further, he thought that there is no real possibility of moral perfection in this life and indeed few of us fully deserve the happiness we are lucky enough to enjoy. Reason cannot prove or disprove the existence of Divine Providence, on Kant's view, nor the immortality of the soul, which seem necessary to rectify these things. Nevertheless, Kant argued, an unlimited amount of time to perfect ourselves (immortality) and a commensurate achievement of wellbeing (insured by God) are "postulates" required by reason when employed in moral matters.

Throughout his moral works, Kant returns time and again to the question of the method moral philosophy should employ when pursuing these aims. A basic theme of these discussions is that the fundamental philosophical issues of morality must be addressed *a priori*, that is, without drawing on observations of human beings and their behavior. Kant's insistence on an *a priori* method to seek out and establish fundamental moral principles, however, does not always appear to be matched by his own practice. The *Metaphysics of Morals*, for instance, is meant to be based on *a priori* rational principles, but many of the specific duties that Kant describes, along with some of the arguments he gives in support of them, rely on general facts about human beings and our circumstances that are known from experience.

In one sense, it might seem obvious why Kant insists on an *a priori* method. A "metaphysics of morals" would be, more or less, an account of the nature and structure of moral requirements — in effect, a categorization of duties and values. Such a project would address such questions as, What *is* a duty? What *kinds* of duties are there? What *is* the good? What *kinds* of goods are there?, and so on. These appear to be metaphysical questions. Any principle used to provide such categorizations appears to be a principle of metaphysics, in a sense, but Kant did not see them as external moral truths that exist independently of rational agents. Moral requirements, instead, are rational principles that tell us what we have overriding reason to do. Metaphysical principles of this sort are always sought out and established by *a priori* methods.

Perhaps something like this was behind Kant's thinking. However, the considerations he offers for an *a priori* method do not all obviously draw on this sort of rationale. The following are three considerations favoring *a priori* methods that he emphasizes repeatedly.

The first is that, as Kant and others have conceived of it, ethics initially requires an analysis of our moral concepts. We must understand the concepts of a "good will", "obligation", "duty" and so on, as well as their logical relationships to one another, before we can determine whether our use of these concepts is justified. Given that the analysis of concepts is an *a priori* matter, to the degree that ethics consists of such an analysis, ethics is *a priori* as well.

Of course, even were we to agree with Kant that ethics should begin with analysis, and that analysis is or should be an entirely *a priori* undertaking, this would not explain why *all* of the fundamental questions of moral philosophy must be pursued *a priori*. Indeed, one of the most important projects of moral philosophy, for Kant, is to show that we, as rational agents, are bound by moral requirements and that fully rational agents would necessarily comply with them. Kant admits that his analytical arguments for the CI are inadequate on their own because the most they can show is that the CI is the supreme principle of morality *if*

there is such a principle. Kant must therefore address the possibility that morality itself is an illusion by showing that the CI really is an unconditional requirement of reason that applies to us. Even though Kant thought that this project of “establishing” the CI must also be carried out *a priori*, he did not think we could pursue this project simply by analyzing our moral concepts or examining the actual behavior of others. What is needed, instead, is a “synthetic”, but still *a priori*, kind of argument that starts from ideas of freedom and rational agency and critically examines the nature and limits of these capacities.

This is the second reason Kant held that fundamental issues in ethics must be addressed with an *a priori* method: The ultimate subject matter of ethics is the nature and content of the principles that necessarily determine a rational will.

Fundamental issues in moral philosophy must also be settled *a priori* because of the nature of moral requirements themselves, or so Kant thought. This is a third reason he gives for an *a priori* method, and it appears to have been of great importance to Kant: Moral requirements present themselves as being *unconditionally necessary*. But an *a posteriori* method seems ill-suited to discovering and establishing what we *must* do whether we feel like doing it or not; surely such a method could only tell us what we *actually* do. So an *a posteriori* method of seeking out and establishing the principle that generates such requirements will not support the presentation of moral “oughts” as unconditional necessities. Kant argued that empirical observations could only deliver conclusions about, for instance, the relative advantages of moral behavior in various circumstances or how pleasing it might be in our own eyes or the eyes of others. Such findings clearly would not support the unconditional necessity of moral requirements. To appeal to *a posteriori* considerations would thus result in a tainted conception of moral requirements. It would view them as demands for which compliance is not unconditionally necessary, but rather necessary only if additional considerations show it to be advantageous, optimistic or in some other way felicitous. Thus, Kant argued that if moral philosophy is to guard against undermining the unconditional necessity of obligation in its analysis and defense of moral thought, it must be carried out entirely *a priori*.

2. Good Will, Moral Worth and Duty

Kant’s analysis of commonsense ideas begins with the thought that the only thing good without qualification is a “good will”. While the phrases “he’s good hearted”, “she’s good natured” and “she means well” are common, “the good will” as Kant thinks of it is not the same as any of these ordinary notions. The idea of a good will is closer to the idea of a “good person”, or, more archaically, a “person of good will”. This use of the term “will” early on in analyzing ordinary moral thought prefigures later and more technical discussions concerning the nature of rational agency. Nevertheless, this idea of a good will is an important commonsense touchstone to which Kant returns throughout his works. The basic idea, as Kant describes it in the *Groundwork*, is that what makes a good person good is his possession of a will that is in a certain way “determined” by, or makes its decisions on the basis of, the moral law. The idea of a good will is supposed to be the idea of one who is committed only to make decisions that she holds to be morally worthy and who takes moral considerations in themselves to be conclusive reasons for guiding her behavior. This sort of disposition or character is something we all highly value, Kant thought. He believes we value it without limitation or qualification. By this, we believe, he means primarily two things.

First, unlike anything else, there is no conceivable circumstance in which we regard our own moral goodness as worth forfeiting simply in order to obtain some desirable object. By contrast, the value of all other desirable qualities, such as courage or cleverness, can be diminished, forgone, or sacrificed under certain circumstances: Courage may be laid aside if it requires injustice, and it is better not to be witty if it requires cruelty. There is no implicit restriction or qualification to the effect that a commitment to give moral considerations decisive weight is worth honoring, *but only under such and such circumstances*.

Second, possessing and maintaining a steadfast commitment to moral principles is the very condition under which anything else is worth having or pursuing. Intelligence and even pleasure are worth having only on the condition that they do not require giving up one’s fundamental moral convictions. The value of a good will thus cannot be that it secures certain valuable ends, whether of our own or of others, since their value is entirely conditional on our possessing and maintaining a good will. Indeed, since a good will is good under any condition, its goodness must not depend on any particular conditions obtaining. Thus, Kant points out that a good will must then also be good *in itself* and not in virtue of its relationship to other things such as the

agent's own happiness, overall welfare or any other effects it may or may not produce. A good will would still "shine like a jewel" even if it were "completely powerless to carry out its aims" (G 4:394).

In Kant's terms, a good will is a will whose decisions are wholly determined by moral demands or, as he often refers to this, by the Moral Law. Human beings inevitably feel this Law as a constraint on their natural desires, which is why such Laws, as applied to human beings, are imperatives and duties. A human will in which the Moral Law is decisive is motivated by the thought of *duty*. A *holy* or *divine* will, if it exists, though good, would not be good because it is motivated by thoughts of duty because such a will does not have natural inclinations and so necessarily fulfills moral requirements without feeling constrained to do so. It is the presence of desires that *could* operate independently of moral demands that makes goodness in human beings a constraint, an essential element of the idea of "duty." So in analyzing unqualified goodness as it occurs in imperfectly rational creatures such as ourselves, we are investigating the idea of being motivated by the thought that we are constrained to act in certain ways that we *might* not want to simply from the thought that we are morally required to do so.

Kant confirms this by comparing motivation by duty with other sorts of motives, in particular, with motives of self-interest, self-preservation, sympathy and happiness. He argues that a dutiful action from any of these motives, however praiseworthy it may be, does not express a good will. Assuming an action has moral worth only if it expresses a good will, such actions have no genuine "moral worth." The conformity of one's action to duty in such cases is only related by accident to morality. For instance, if one is motivated by happiness alone, then had conditions not conspired to align one's duty with one's own happiness one would not have done one's duty. By contrast, were one to supplant any of these motivations with the motive of duty, the morality of the action would then express one's determination to act dutifully out of respect for the moral law itself. Only then would the action have moral worth.

Kant's views in this regard have understandably been the subject of much controversy. Many object that we do not think better of actions done for the sake of duty than actions performed out of emotional concern or sympathy for others, especially those things we do for friends and family. Worse, moral worth appears to require not only that one's actions be motivated by duty, but also that no other motives, even love or friendship, cooperate. Yet Kant's defenders have argued that his point is not that we do not admire or praise motivating concerns other than duty, only that from the point of view of someone deliberating about what to do, these concerns are not decisive in the way that considerations of moral duty are. What is crucial in actions that express a good will is that in conforming to duty a perfectly virtuous person always would, and so ideally we should, recognize and be moved by the thought that our conformity is morally obligatory. The motivational structure of the agent should be arranged so that she always treats considerations of duty as sufficient reasons for conforming to those requirements. In other words, we should have a firm commitment not to perform an action if it is morally forbidden and to perform an action if it is morally required. Having a good will, in this sense, is compatible with having feelings and emotions of various kinds, and even with aiming to cultivate some of them in order to counteract desires and inclinations that tempt us to immorality. Controversy persists, however, about whether Kant's claims about the motive of duty go beyond this basic point (Timmermann 2007; Herman 1993; Wood 1998; Baron 1995).

Suppose for the sake of argument we agree with Kant. We now need to know what distinguishes the principle that lays down our duties from these other motivating principles, and so makes motivation by it the source of unqualified value.

3. Duty and Respect for Moral Law

According to Kant, what is singular about motivation by duty is that it consists of bare respect for the moral law. What naturally comes to mind is this: Duties are rules or laws of some sort combined with some sort of felt constraint or incentive on our choices, whether from external coercion by others or from our own powers of reason. For instance, the bylaws of a club lay down duties for its officers and enforce them with sanctions. City and state laws establish the duties of citizens and enforce them with coercive legal power. Thus, if we do something because it is our "civic" duty, or our duty "as a boy scout" or "a good American," our motivation is respect for the code that makes it our duty. Thinking we are duty bound is simply respecting, as such, certain laws pertaining to us.

However intuitive, this cannot be all of Kant's meaning. For one thing, as with the Jim Crow laws of the old South and the Nuremberg laws of Nazi Germany, the laws to which these types of "actions from duty" conform may be morally despicable. Respect for such laws could hardly be thought valuable. For another, our motive in conforming our actions to civic and other laws is rarely unconditional respect. We also have an eye toward doing our part in maintaining civil or social order, toward punishments or loss of standing and reputation in violating such laws, and other outcomes of lawful behavior. Indeed, we respect these laws to the degree, but only to the degree, that they do not violate values, laws or principles we hold more dear. Yet Kant thinks that, in acting from duty, we are not at all motivated by a prospective outcome or some other extrinsic feature of our conduct except insofar as these are requirements of duty itself. We are motivated by the mere conformity of our will to law as such.

To act out of respect for the moral law, in Kant's view, is to be moved to act by a recognition that the moral law is a supremely authoritative standard that binds us and to experience a kind of feeling, which is akin to awe and fear, when we acknowledge the moral law as the source of moral requirements. Human persons inevitably have respect for the moral law even though we are not always moved by it and even though we do not always comply with the moral standards that we nonetheless recognize as authoritative.

Kant's account of the content of moral requirements and the nature of moral reasoning is based on his analysis of the unique force moral considerations have as reasons to act. The force of moral requirements as reasons is that we cannot ignore them no matter how circumstances might conspire against any other consideration. Basic moral requirements retain their reason-giving force under any circumstance, they have universal validity. So, whatever else may be said of basic moral requirements, their content is universal. Only a universal law could be the content of a requirement that has the reason-giving force of morality. This brings Kant to a preliminary formulation of the CI: "I ought never to act except in such a way that I could also will that my maxim should become a universal law" (G 4:402). This is the principle which motivates a good will, and which Kant holds to be the fundamental principle of all of morality.

4. Categorical and Hypothetical Imperatives

Kant holds that the fundamental principle of our moral duties is a *categorical imperative*. It is an *imperative* because it is a command addressed to agents who could follow it but might not (e.g., "Leave the gun. Take the cannoli."). It is *categorical* in virtue of applying to us unconditionally, or simply because we possess rational wills, without reference to any ends that we might or might not have. It does not, in other words, apply to us on the condition that we have antecedently adopted some goal for ourselves.

There are "oughts" other than our moral duties, according to Kant, but these oughts are distinguished from the moral ought in being based on a quite different kind of principle, one that is the source of *hypothetical imperatives*. A hypothetical imperative is a command that also applies to us in virtue of our having a rational will, but not *simply* in virtue of this. It requires us to exercise our wills in a certain way *given* we have antecedently willed an end. A hypothetical imperative is thus a command in a conditional form. But not any command in this form counts as a hypothetical imperative in Kant's sense. For instance, "if you're happy and you know it, clap your hands!" is a conditional command. But the antecedent conditions under which the command "clap your hands" applies to you do not posit any end that you will, but consist rather of emotional and cognitive states you may or may not be in. Further, "if you want pastrami, try the corner deli" is also a command in conditional form, but strictly speaking it too fails to be a hypothetical imperative in Kant's sense since this command does not apply to us in virtue of our *willing* some end, but only in virtue of our *desiring* or *wanting* an end. For Kant, willing an end involves more than desiring; it requires actively choosing or committing to the end rather than merely finding oneself with a passive desire for it. Further, there is nothing irrational in failing to will means to what one desires. An imperative that applied to us in virtue of our desiring some end would thus not be a hypothetical imperative of *practical rationality* in Kant's sense.

The condition under which a hypothetical imperative applies to us, then, is that we will some end. Now, for the most part, the ends we will we might not have willed, and some ends that we do not will we might nevertheless have willed. But there is at least conceptual room for the idea of a natural or inclination-based end that we *must* will. The distinction between ends that we might or might not will and those, if any, we necessarily will as the kinds of natural beings we are, is the basis for his distinction between two kinds of hypothetical imperatives. Kant names these "problematic" and "assertoric", based on how the end is willed.

If the end is one that we might or might not will — that is, it is a merely *possible* end — the imperative is problematic. For instance, “Don’t ever take side with anyone against the Family.” is a problematic imperative, even if the end posited here is (apparently) one’s own continued existence. Almost all non-moral, rational imperatives are problematic, since there are virtually no ends that we necessarily will as human beings.

As it turns out, the only (non-moral) end that we will, as a matter of natural necessity, is our own happiness. Any imperative that applied to us because we will our own happiness would thus be an *assertoric* imperative. Rationality, Kant thinks, can issue no imperative if the end is indeterminate, and happiness is an indeterminate end. Although we can say for the most part that if one is to be happy, one should save for the future, take care of one’s health and nourish one’s relationships, these fail to be genuine commands in the strictest sense and so are instead mere “counsels.” Some people are happy without these, and whether you could be happy without them is, although doubtful, an open question.

Since Kant presents moral and prudential rational requirements as first and foremost demands on our wills rather than on external acts, moral and prudential evaluation is first and foremost an evaluation of the will our actions express. Thus, it is not an error of rationality to fail to take the necessary means to one’s (willed) ends, nor to fail to *want* to take the means; one only falls foul of non-moral practical reason if one fails to *will* the means. Likewise, while actions, feelings or desires may be the focus of other moral views, for Kant practical irrationality, both moral and prudential, focuses mainly on our willing.

One recent interpretive dispute (Hill 1973; Schroeder 2009; Rippon 2014) has been about whether hypothetical imperatives, in Kant’s view, have a “wide” or “narrow” scope. That is, do such imperatives tell us to take the necessary means to our ends or give up our ends (wide scope) or do they simply tell us that, if we have an end, then take the necessary means to it.

Kant describes the will as operating on the basis of subjective volitional principles he calls “maxims”. Hence, morality and other rational requirements are, for the most part, demands that apply to the maxims that we act on. . The form of a maxim is “I will *A* in *C* in order to realize or produce *E*” where “*A*” is some act type, “*C*” is some type of circumstance, and “*E*” is some type of end to be realized or achieved by *A* in *C*. Since this is a principle stating only what some *agent* wills, it is *subjective*. (A principle that governs *any* rational will is an *objective* principle of volition, which Kant refers to as a practical law). For anything to count as human willing, it must be based on a maxim to pursue some end through some means. Hence, in employing a maxim, any human willing already embodies the form of means-end reasoning that calls for evaluation in terms of hypothetical imperatives. To that extent at least, then, *anything* dignified as human willing is subject to rational requirements.

5. The Formula of the Universal Law of Nature

Kant’s first formulation of the CI states that you are to “act only in accordance with that maxim through which you can at the same time will that it become a universal law” (G 4:421). O’Neill (1975, 1989) and Rawls (1980, 1989), among others, take this formulation in effect to summarize a decision procedure for moral reasoning, and we will follow their basic outline: First, formulate a maxim that enshrines your reason for acting as you propose. Second, recast that maxim as a universal law of nature governing all rational agents, and so as holding that all must, by natural law, act as you yourself propose to act in these circumstances. Third, consider whether your maxim is even conceivable in a world governed by this law of nature. If it is, then, fourth, ask yourself whether you would, or could, rationally *will* to act on your maxim in such a world. If you could, then your action is morally permissible.

If your maxim fails the third step, you have a “perfect” duty admitting “of no exception in favor of inclination” to refrain from acting on that maxim (G 4:421). If your maxim fails the fourth step, you have an “imperfect” duty requiring you to pursue a policy that can admit of such exceptions. If your maxim passes all four steps, only then is acting on it morally permissible. Following Hill (1971), we can understand the difference in duties as formal: Perfect duties come in the form “One must never (or always) ϕ to the fullest extent possible in *C*”, while imperfect duties, since they require us to adopt an end, at least require that “One must sometimes and to some extent ϕ in *C*.” So, for instance, Kant held that the maxim of committing suicide to avoid future unhappiness did not pass the third step, the contradiction in conception test. Hence,

one is forbidden to act on the maxim of committing suicide to avoid unhappiness. By contrast, the maxim of refusing to assist others in pursuit of their projects passes the contradiction in conception test, but fails the contradiction in the will test at the fourth step. Hence, we have a duty to sometimes and to some extent aid and assist others.

Kant held that ordinary moral thought recognized moral duties toward ourselves as well as toward others. Hence, together with the distinction between perfect and imperfect duties, Kant recognized four categories of duties: perfect duties toward ourselves, perfect duties toward others, imperfect duties toward ourselves and imperfect duties toward others. Kant uses four examples in the *Groundwork*, one of each kind of duty, to demonstrate that every kind of duty can be derived from the CI, and hence to bolster his case that the CI is indeed the fundamental principle of morality. To refrain from suicide is a perfect duty toward oneself; to refrain from making promises you have no intention of keeping is a perfect duty toward others; to develop one's talents is an imperfect duty toward oneself; and to contribute to the happiness of others is an imperfect duty toward others. Again, Kant's interpreters differ over exactly how to reconstruct the derivation of these duties. We will briefly sketch one way of doing so for the perfect duty to others to refrain from lying promises and the imperfect duty to ourselves to develop talents.

Kant's example of a perfect duty to others concerns a promise you might consider making but have no intention of keeping in order to get needed money. Naturally, being rational requires not contradicting oneself, but there is no self-contradiction in the maxim "I will make lying promises when it achieves something I want." An immoral action clearly does not involve a self-contradiction in this sense (as would the maxim of finding a married bachelor). Kant's position is that it is irrational to perform an action if that action's maxim contradicts itself *once made into a universal law of nature*. The maxim of lying whenever it gets you what you want generates a contradiction once you try to combine it with the universalized version that all rational agents must, by a law of nature, lie when doing so gets them what they want.

Here is one way of seeing how this might work: If I conceive of a world in which everyone by nature must try to deceive people any time this will get them what they want, I am conceiving of a world in which no practice of giving one's word could ever arise and, because this is a law of nature, we can assume that it is widely known that no such practice could exist. So I am conceiving of a world in which everyone knows that no practice of giving one's word exists. My maxim, however, is to make a deceptive promise in order to get needed money. And it is a necessary means of doing this that a practice of taking the word of others exists, so that someone might take my word and I take advantage of their doing so. Thus, in trying to conceive of my maxim in a world in which no one ever takes anyone's word in such circumstances, and knows this about one another, I am trying to conceive of this: A world in which no practice of giving one's word exists, but also, at the very same time, a world in which just such a practice does exist, for me to make use of in my maxim. It is a world containing my promise and a world in which there can be no promises. Hence, it is inconceivable that I could sincerely act on my maxim in a world in which my maxim is a universal law of nature. Since it is inconceivable that these two things could exist together, I am forbidden ever to act on the maxim of lying to get money.

By contrast with the maxim of the lying promise, we can easily conceive of adopting a maxim of refusing to develop any of our talents in a world in which that maxim is a universal law of nature. It would undoubtedly be a world more primitive than our own, but pursuing such a policy is still conceivable in it. However, it is not, Kant argues, possible to rationally will this maxim in such a world. The argument for why this is so, however, is not obvious, and some of Kant's thinking seems hardly convincing: Insofar as we are rational, he says, we already necessarily will that all of our talents and abilities be developed. Hence, although I can conceive of a talentless world, I cannot rationally will that it come about, given that I already will, insofar as I am rational, that I develop all of my own. Yet, given limitations on our time, energy and interest, it is difficult to see how full rationality requires us to aim to fully develop literally all of our talents. Indeed, it seems to require much less, a judicious picking and choosing among one's abilities. Further, all that is required to show that I cannot will a talentless world is that, insofar as I am rational, I necessarily will that *some* talents in me be developed, not the dubious claim that I rationally will that they *all* be developed. Moreover, suppose rationality did require me to aim at developing all of my talents. Then, there seems to be no need to go further in the CI procedure to show that refusing to develop talents is immoral. Given that, insofar as we are rational, we must will to develop capacities, it is by this very fact irrational not to do so.

However, mere failure to conform to something we rationally will is not yet immorality. Failure to conform to instrumental principles, for instance, is irrational but not always immoral. In order to show that this maxim is categorically forbidden, one strategy is to make use of several other of Kant's claims or assumptions.

First, we must accept Kant's claim that, by "natural necessity," we will our own happiness as an end (G 4:415). This is a claim he uses not only to distinguish assertoric from problematic imperatives, but also to argue for the imperfect duty of helping others (G 4:423) He also appears to rely on this claim in each of his examples. Each maxim he is testing appears to have happiness as its aim. One explanation for this is that, since each person necessarily wills her own happiness, maxims in pursuit of this goal will be the typical object of moral evaluation. This, at any rate, is clear in the talents example itself: The forbidden maxim adopted by the ne'er-do-well is supposed to be "devoting his life solely to...enjoyment" (G 4:423) rather than to developing his talents.

Second, we must assume, as also seems reasonable, that a necessary means to achieving (normal) human happiness is not only that we ourselves develop some talent, but also that others develop some capacities of theirs at some time. For instance, I cannot engage in the normal pursuits that make up my own happiness, such as playing piano, writing philosophy or eating delicious meals, unless I have developed some talents myself, and, moreover, someone else has made pianos and written music, taught me writing, harvested foods and developed traditions of their preparation.

Finally, Kant's examples come on the heels of defending the position that rationality requires conformity to hypothetical imperatives. Thus, we should assume that, necessarily, rational agents will the necessary and available means to any ends that they will. And once we add this to the assumptions that we must will our own happiness as an end, and that developed talents are necessary means to achieving that end, it follows that we cannot rationally will that a world come about in which it is a law that no one ever develops any of their natural talents. We cannot do so, because our own happiness is the very end contained in the maxim of giving ourselves over to pleasure rather than self-development. Since we will the necessary and available means to our ends, we are rationally committed to willing that everyone sometime develop his or her talents. So since we cannot will as a universal law of nature that no one ever develop any talents — given that it is inconsistent with what we now see that we rationally will — we are forbidden from adopting the maxim of refusing to develop any of our own.

6. The Humanity Formula

Most philosophers who find Kant's views attractive find them so because of the Humanity Formulation of the CI. This formulation states that we should never act in such a way that we treat humanity, whether in ourselves or in others, as a means only but always as an end in itself. This is often seen as introducing the idea of "respect" for persons, for whatever it is that is essential to our humanity. Kant was clearly right that this and the other formulations bring the CI "closer to intuition" than the Universal Law formula. Intuitively, there seems something wrong with treating human beings as mere instruments with no value beyond this. But this very intuitiveness can also invite misunderstandings.

First, the Humanity Formula does not rule out using people as means to our ends. Clearly this would be an absurd demand, since we apparently do this all the time in morally appropriate ways. Indeed, it is hard to imagine any life that is recognizably human without the use of others in pursuit of our goals. The food we eat, the clothes we wear, the chairs we sit on and the computers we type at are gotten only by way of talents and abilities that have been developed through the exercise of the wills of many people. What the Humanity Formula rules out is engaging in this pervasive use of humanity in such a way that we treat it as a *mere* means to our ends. Thus, the difference between a horse and a taxi driver is not that we may use one but not the other as a means of transportation. Unlike a horse, the taxi driver's humanity must at the same time be treated as an end in itself.

Second, it is not human beings *per se* but the "humanity" in human beings that we must treat as an end in itself. Our "humanity" is that collection of features that make us distinctively human, and these include capacities to engage in self-directed rational behavior and to adopt and pursue our own ends, and any other rational capacities necessarily connected with these. Thus, supposing that the taxi driver has freely exercised his rational capacities in pursuing his line of work, we make permissible use of these capacities as a means

only if we behave in a way that he could, when exercising his rational capacities, consent to — for instance, by paying an agreed on price.

Third, the idea of an end has three senses for Kant, two positive senses and a negative sense. An end in the first positive sense is a thing we will to produce or bring about in the world. For instance, if losing weight is my end, then losing weight is something I aim to bring about. An end in this sense guides my actions in that once I will to produce something, I then deliberate about and aim to pursue means of producing it if I am rational. Humanity is not an “end” in this sense, though even in this case, the end “lays down a law” for me. Once I have adopted an end in this sense, it dictates that I do something: I should act in ways that will bring about the end or instead choose to abandon my goal.

An end in the negative sense lays down a law for me as well, and so guides action, but in a different way. Korsgaard (1996) offers self-preservation as an example of an end in a negative sense: We do not try to produce our self-preservation. Rather, the end of self-preservation prevents us from engaging in certain kinds of activities, for instance, picking fights with mobsters, and so on. That is, as an end, it is something I do not act *against* in pursuing my positive ends, rather than something I produce.

Humanity is in the first instance an end in this negative sense: It is something that *limits* what I may do in pursuit of my other ends, similar to the way that my end of self-preservation limits what I may do in pursuit of other ends. Insofar as it *limits* my actions, it is a source of *perfect* duties. Now many of our ends are subjective in that they are not ends that every rational being must have. Humanity is an *objective* end, because it is an end that every rational being must have. Hence, my own humanity as well as the humanity of others limit what I am *morally permitted* to do when I pursue my other, non-mandatory, ends.

The humanity in myself and others is also a *positive* end, though not in the first positive sense above, as something to be produced by my actions. Rather, it is something to realize, cultivate or further by my actions. Becoming a philosopher, pianist or novelist might be my end in this sense. When my end is becoming a pianist, my actions do not, or at least not simply, produce something, being a pianist, but constitute or realize the activity of being a pianist. Insofar as the humanity in ourselves must be treated as an end in itself in this second *positive* sense, it must be cultivated, developed or fully actualized. Hence, the humanity in oneself is the source of a duty to develop one’s talents or to “perfect” one’s humanity. When one makes one’s own humanity one’s end, one pursues its development, much as when one makes becoming a pianist one’s end, one pursues the development of piano playing. And insofar as humanity is a positive end in others, I must attempt to further their ends as well. In so doing, I further the humanity in others, by helping further the projects and ends that they have willingly adopted for themselves. It is this sense of humanity as an end-in-itself on which some of Kant’s arguments for imperfect duties rely.

Finally, Kant’s Humanity Formula requires “respect” for the humanity in persons. Proper regard for something with absolute value or worth requires respect for it. But this can invite misunderstandings. One way in which we respect persons, termed “appraisal respect” by Stephen Darwall (1977), is clearly not the same as the kind of respect required by the Humanity Formula: I may respect you as a rebounder but not a scorer, or as a researcher but not as a teacher. When I respect you in this way, I am positively appraising you in light of some achievement or virtue you possess relative to some standard of success. If this were the sort of respect Kant is counseling then clearly it may vary from person to person and is surely not what treating something as an end-in-itself requires. For instance, it does not seem to prevent me from regarding rationality as an achievement and respecting one person as a rational agent in this sense, but not another. And Kant is not telling us to ignore differences, to pretend that we are blind to them on mindless egalitarian grounds. However, a distinct way in which we respect persons, referred to as “recognition respect” by Darwall, better captures Kant’s position: I may respect you because you are a student, a Dean, a doctor or a mother. In such cases of respecting you because of who or what you are, I am giving the proper regard to a certain fact about you, your being a Dean for instance. This sort of respect, unlike appraisal respect, is not a matter of degree based on your having measured up to some standard of assessment. Respect for the humanity in persons is more like Darwall’s recognition respect. We are to respect human beings simply because they are persons and this requires a certain sort of regard. We are not called on to respect them insofar as they have met some standard of evaluation appropriate to persons. And, crucially for Kant, persons cannot lose their humanity by their misdeeds – even the most vicious persons, Kant thought, deserve basic respect as persons with humanity.

7. The Autonomy Formula

The third formulation of the CI is “the Idea of the will of every rational being as *a will that legislates universal law*.” (G 4:432). Although Kant does not state this as an imperative, as he does in the other formulations, it is easy enough to put it in that form: Act so that through your maxims you could be a legislator of universal laws. This sounds very similar to the first formulation. However, in this case we focus on our status as universal law *givers* rather than universal law *followers*. This is of course the source of the very dignity of humanity Kant speaks of in the second formulation. A rational will that is merely bound by universal laws could act accordingly from natural and non-moral motives, such as self-interest. But in order to be a legislator of universal laws, such contingent motives, motives that rational agents such as ourselves may or may not have, must be set aside. Hence, we are required, according to this formulation, to conform our behavior to principles that express this autonomy of the rational will — its status as a source of the very universal laws that obligate it. As with the Humanity Formula, this new formulation of the CI does not change the outcome, since each is supposed to formulate the very same moral law, and in some sense “unite” the other formulations within it. Kant takes each formulation that succeeds the first in its own way as bringing the moral law “closer to feeling”. The Autonomy Formula presumably does this by putting on display the source of our dignity and worth, our status as free rational agents who are the source of the authority behind the very moral laws that bind us.

8. The Kingdom of Ends Formula

This formulation has gained favor among Kantians in recent years (see Rawls, 1972; Hill, 1972). Many see it as introducing more of a social dimension to Kantian morality. Kant states that the above concept of every rational will as a will that must regard itself as enacting laws binding all rational wills is closely connected to another concept, that of a “systematic union of different rational beings under common laws”, or a “Kingdom of Ends” (G 4:433). The formulation of the CI states that we must “act in accordance with the maxims of a member giving universal laws for a merely possible kingdom of ends” (G 4:439). It combines the others in that (i) it requires that we conform our actions to the laws of an ideal moral legislature, (ii) that this legislature lays down universal laws, binding all rational wills including our own, and (iii) that those laws are of “a merely possible kingdom” each of whose members equally possesses this status as legislator of universal laws, and hence must be treated always as an end in itself. The intuitive idea behind this formulation is that our fundamental moral obligation is to act only on principles which could earn acceptance by a community of fully rational agents each of whom have an equal share in legislating these principles for their community.

9. The Unity of the Formulas

Kant claimed that all of these CI formulas were equivalent. Unfortunately, he does not say in what sense. What he *says* is that these “are basically only so many formulations of precisely the same law, each one of them by itself uniting the other two within it,” and that the differences between them are “more subjectively than objectively practical” in the sense that each aims “to bring an Idea of reason closer to intuition (by means of a certain analogy) and thus nearer to feeling” (G 4:435). He also says that one formula “follows from” another (G 4:431), and that the concept foundational to one formula “leads to a closely connected” concept at the basis of another formula (G 4:433). Thus, his claim that the formulations are equivalent could be interpreted in a number of ways.

Kant’s statement that each formula “unites the other two within it” initially suggests that the formulas are equivalent in *meaning*, or at least one could analytically derive one formula from another. Some of Kant’s commentators, for example, have argued along the following lines: That I should always treat humanity as an end in itself entails that I should act only on maxims that are consistent with themselves as universal laws of nature (O’Neill 1975, 1990; Engstrom 2009; Sensen 2011). There are remaining doubts some commentators have, however, about whether this strategy can capture the full meaning of the Humanity Formula or explain all of the duties that Kant claims to derive from it (Wood 1999, 2007; Cureton 2013).

Perhaps, then, if the formulas are not equivalent in meaning, they are nevertheless logically interderivable and hence equivalent in this sense. The universal law formula is not itself derived, as some of Kant’s

interpreters have suggested, from the principle of non-contradiction. That would have the consequence that the CI is a logical truth, and Kant insists that it is not or at least that it is not analytic. Since the CI formulas are not logical truths, then, it is possible that they could be logically interderivable. However, despite his claim that each contains the others within it, what we find in the *Groundwork* seems best interpreted as a derivation of each successive formula from the immediately preceding formula. There are, nonetheless, a few places in which it seems that Kant is trying to work in the opposite direction. One is found in his discussion of the Humanity Formula. There Kant says that only something “*whose existence in itself* had an absolute worth” could be the ground of a categorically binding law (G 4:428). He then boldly proclaims that humanity is this absolutely valuable thing, referring to this as a “postulate” that he will argue for in the final chapter of the *Groundwork* (G 4:429n). One might take this as expressing Kant’s intention to derive thereby the universal law formula from the Humanity Formula: *If* something is absolutely valuable, *then* we must act only on maxims that can be universal laws. But (he postulates) humanity is absolutely valuable. *Thus*, we must act only on maxims that can be universal laws. This (we think) anomolous discussion may well get at some deep sense in which Kant thought the formulations were equivalent. Nonetheless, this derivation of the universal law formulation from the Humanity Formulation seems to require a substantive, *synthetic* claim, namely, that humanity is indeed absolutely valuable. And if it does require this, then, contrary to Kant’s own insistence, the argument of *Groundwork II* does not appear to be merely an analytic argument meant simply to establish the content of the moral law.

The most straightforward interpretation of the claim that the formulas are equivalent is as the claim that following or applying each formula would generate all and only the same duties (Allison 2011). This seems to be supported by the fact that Kant used the same examples through the Law of Nature Formula and the Humanity Formula. Thus, the Universal Law Formulation generates a duty to ϕ if and only if the Humanity Formula generates a duty to ϕ , (and so on for the other formulations). In other words, respect for humanity as an end in itself could never lead you to act on maxims that would generate a contradiction when universalized, and vice versa. This way of understanding Kant’s claim also fits with his statement that there is no “objective practical difference” between the formulations although there are “subjective” differences. The subjective differences between formulas are presumably differences that appeal in different ways to various conceptions of what morality demands of us. But this difference in meaning is compatible with there being no practical difference, in the sense that conformity to one formulation cannot lead one to violate another formulation.

10. Autonomy

At the heart of Kant’s moral theory is the idea of autonomy. Most readers interpret Kant as holding that autonomy is a property of rational wills or agents. Understanding the idea of autonomy was, in Kant’s view, key to understanding and justifying the authority that moral requirements have over us. As with Rousseau, whose views influenced Kant, freedom does not consist in being bound by no law, but by laws that are in some sense of one’s own making. The idea of freedom as autonomy thus goes beyond the merely “negative” sense of being *free from* causes on our conduct originating outside of ourselves. It contains first and foremost the idea of laws made and laid down by oneself, and, in virtue of this, laws that have decisive authority over oneself.

Kant’s basic idea can be grasped intuitively by analogy with the idea of political freedom as autonomy (See Reath 1994). Consider how political freedom in liberal theories is thought to be related to legitimate political authority: A state is free when its citizens are bound only by laws in some sense of their own making — created and put into effect, say, by vote or by elected representatives. The laws of that state then express the will of the citizens who are bound by them. The idea, then, is that the source of legitimate political authority is not external to its citizens, but internal to them, internal to “the will of the people.” It is because the body politic created and enacted these laws for itself that it can be bound by them. An autonomous state is thus one in which the authority of its laws is in the will of the people in that state, rather than in the will of a people external to that state, as when one state imposes laws on another during occupation or colonization. In the latter case, the laws have no legitimate authority over those citizens. In a similar fashion, we may think of a person as free when bound only by her own will and not by the will of another. Her actions then express her own will and not the will of someone or something else. The authority of the principles binding her will is then also not external to her will. It comes from the fact that she willed them. So autonomy, when applied to an individual, ensures that the source of the authority of the principles that bind her is in her own will.

Kant's view can be seen as the view that the moral law is just such a principle. Hence, the "moral legitimacy" of the CI is grounded in its being an expression of each person's own rational will. It is because each person's own reason is the legislator and executor of the moral law that it is authoritative for her. (For a contrasting interpretation of autonomy that emphasizes the intrinsic value of freedom of choice and the instrumental role of reason in preserving that value, see Guyer 2007).

Kant argues that the idea of an autonomous will emerges from a consideration of the idea of a will that is free "in a negative sense." The concept of a rational will is of a will that operates by responding to what it takes to be reasons. This is, firstly, the concept of a will that does not operate through the influence of factors outside of this responsiveness to apparent reasons. For a will to be free is thus for it to be physically and psychologically unforced in its operation. Hence, behaviors that are performed because of obsessions or thought disorders are not free in this negative sense. But also, for Kant, a will that operates by being determined through the operation of natural laws, such as those of biology or psychology, cannot be thought of as operating by responding to reasons. Hence, determination by natural laws is conceptually incompatible with being free in a negative sense.

A crucial move in Kant's argument is his claim that a rational will cannot act except "under the Idea" of its own freedom (G 4:448). The expression "acting under the Idea of freedom" is easy to misunderstand. It does not mean that a rational will must *believe* it is free, since determinists are as free as libertarians in Kant's view. Indeed, Kant goes out of his way in his most famous work, the *Critique of Pure Reason*, to argue that we have no rational basis for *believing* our wills to be free. This would involve, he argues, attributing a property to our wills that they would have to have as 'things in themselves' apart from the causally determined world of appearances. Of such things, he insists, we can have no knowledge. For much the same reason, Kant is not claiming that a rational will cannot operate without *feeling* free. Feelings, even the feeling of operating freely or the "looseness" Hume refers to when we act, cannot be used in an *a priori* argument to establish the CI, since they are empirical data.

One helpful way to understand acting "under the Idea of freedom" is by analogy with acting "under the Idea" that there are purposes in nature: Although there is, according to Kant, no rational basis for the belief that the natural world is (or is not) arranged according to some purpose by a Designer, the actual practices of science often require looking for the purpose of this or that chemical, organ, creature, environment, and so on. Thus, one engages in these natural sciences by searching for purposes in nature. Yet when an evolutionary biologist, for instance, looks for the purpose of some organ in some creature, she does not after all thereby believe that the creature *was* designed that way, for instance, by a Deity. Nor is she having some feeling of "designedness" in the creature. To say that she "acts under the Idea of" design is to say something about the practice of biology: Practicing biology involves searching for the purposes of the parts of living organisms. In much the same way, although there is no rational justification for the belief that our wills are (or are not) free, the actual practice of practical deliberation and decision consists of a search for the right causal chain of which to be the origin — consists, that is, seeking to be the first causes of things, wholly and completely through the exercise of one's own will.

Kant says that a will that cannot exercise itself except under the Idea of its freedom *is* free from a practical point of view (*im practischer Absicht*). In saying such wills are free from a practical point of view, he is saying that in engaging in practical endeavors — trying to decide what to do, what to hold oneself and others responsible for, and so on — one is justified in holding oneself to all of the principles to which one would be justified in holding wills that are autonomous free wills. Thus, once we have established the set of prescriptions, rules, laws and directives that would bind an autonomous free will, we then hold ourselves to this very same set of prescriptions, rules, laws and directives. And one is justified in this because rational agency can only operate by seeking to be the first cause of its actions, and these are the prescriptions, and so on, of being a first cause of action. Therefore, rational agents are free in a negative sense insofar as any practical matter is at issue.

Crucially, rational wills that are negatively free must be autonomous, or so Kant argues. This is because the will is a kind of cause—willing causes action. Kant took from Hume the idea that causation implies universal regularities: if x causes y , then there is some universally valid law connecting X s to Y s. So, if my will is the cause of my ϕ ing, then Φ ing is connected to the sort of willing I engage in by some universal law. But it can't be a natural law, such as a psychological, physical, chemical or biological law. These laws, which Kant thought were universal too, govern the movements of my body, the workings of my brain and nervous

system and the operation of my environment and its effects on me as a material being. But they cannot be the laws governing the operation of my will; that, Kant already argued, is inconsistent with the freedom of my will in a negative sense. So, the will operates according to a universal law, though not one authored by nature, but one of which I am the origin or author. And that is to say that, in viewing my willing to ϕ as a negatively free cause of my ϕ ing, I must view my will as the autonomous cause of my having ϕ ed, as causing my having ϕ ed by way of some law that I, insofar as I am a rational will, laid down for my will.

Thus, Kant argues, a rational will, insofar as it is rational, is a will conforming itself to those laws valid for any rational will. Addressed to imperfectly rational wills, such as our own, this becomes an imperative: "Conform your action to a universal non-natural law." Kant assumed that there was some connection between this formal requirement and the formulation of the CI which enjoins us to "Act as though the maxim of your action were to become by your will a universal law of nature." But, as commentators have long noticed (see, e.g., Hill, 1989a, 1989b), it is not clear what the link is between the claim that rational autonomous wills conform themselves to whatever universally valid laws require, and the more substantial and controversial claim that you should evaluate your maxims in the ways implied by the universal law of nature formulation.

Kant appeared not to recognize the gap between the law of an autonomous rational will and the CI, but he was apparently unsatisfied with the argument establishing the CI in *Groundwork* III for another reason, namely, the fact that it does not prove that we really are free. In the *Critique of Practical Reason*, he states that it is simply a "fact of reason" (*Factum der Vernunft*) that our wills are bound by the CI, and he uses this to argue that our wills are autonomous. Hence, while in the *Groundwork* Kant relies on a dubious argument for our autonomy to establish that we are bound by the moral law, in the second *Critique*, he argues from the bold assertion of our being bound by the moral law to our autonomy.

The apparent failure of Kant's argument to establish the autonomy of the will, and hence the authority of moral demands over us, has not deterred his followers from trying to make good on this project. One strategy favored recently has been to turn back to the arguments of *Groundwork II* for help. Kant himself repeatedly claimed that these arguments are merely analytic but that they do not establish that there is anything that answers to the concepts he analyzes. The conclusions are thus fully compatible with morality being, as he puts it, a "mere phantom of the brain" (G 4:445). Kant clearly takes himself to have established that rational agents such as ourselves must take the means to our ends, since this is analytic of rational agency. But there is a chasm between this analytic claim and the supposed synthetic conclusion that rational agency also requires conforming to a further, non-desire based, principle of practical reason such as the CI. Nevertheless, some see arguments in *Groundwork II* that establish just this. These strategies involve a new "teleological" reading of Kant's ethics that relies on establishing the existence of an absolute value or an "end in itself" (We say more about this teleological reading below.). They begin with Kant's own stated assumption that there is such an end in itself if and only if there is a categorical imperative binding on all rational agents as such. If this assumption is true, then if one can on independent grounds prove that there is something which is an end in itself, one will have an argument for a categorical imperative. One such strategy, favored by Korsgaard (1996) and Wood (1999) relies on the apparent argument Kant gives that humanity is an end in itself. Guyer, by contrast, sees an argument for freedom as an end in itself (Guyer 2000). Both strategies have faced textual and philosophical hurdles. Considerable interpretive finesse, for instance, is required to explain Kant's stark insistence on the priority of principles and law over the good in the second *Critique* (CPrR 5:57–67)

Although most of Kant's readers understand the property of autonomy as being a property of rational wills, some, such as Thomas E. Hill, have held that Kant's central idea is that of autonomy is a property, not primarily of wills, but of principles. The core idea is that Kant believed that all moral theories prior to his own went astray because they portrayed fundamental moral principles as appealing to the existing interests of those bound by them. By contrast, in Kant's view moral principles must not appeal to such interests, for no interest is necessarily universal. Thus, in assuming at the outset that moral principles must embody some interest (or "heteronomous" principles), such theories rule out the very possibility that morality is universally binding. By contrast, the Categorical Imperative, because it does not enshrine existing interests, presumes that rational agents can conform to a principle that does not appeal to their interests (or an "autonomous" principle), and so can fully ground our conception, according to Kant, of what morality requires of us.

A different interpretive strategy, which has gained prominence in recent years, focuses on Kant's apparent identification, in *Groundwork* III, of the will and practical reason. One natural way of interpreting Kant's

conception of freedom is to understand it in terms of the freedom and spontaneity of reason itself. This in turn apparently implies that our wills are necessarily aimed at what is rational and reasonable. To will something, on this picture, is to govern oneself in accordance with reason. Often, however, we fail to effectively so govern ourselves because we are imperfect rational beings who are caused to act by our non-rational desires and inclinations. The result, at least on one version of this interpretation (Wolff 1973), is that we either act rationally and reasonably (and so autonomously) or we are merely caused to behave in certain ways by non-rational forces acting on us (and so heteronomously). This is, however, an implausible view. It implies that all irrational acts, and hence all immoral acts, are not willed and therefore not free. Most interpreters have denied that this is the proper interpretation of Kant's views. However, several prominent commentators nonetheless think that there is some truth in it (Engstrom 2009; Reath 2015; Korsgaard 1996, 2008, 2009). They agree that we always act under the "guise of the good" in the sense that our will is necessarily aimed at what is objectively and subjectively rational and reasonable, but these interpreters also think that, for Kant, there is a middle-ground between perfect conformity to reason and being caused to act by natural forces. In particular, when we act immorally, we are either weak-willed or we are misusing our practical reason by willing badly. We do not have the capacity to aim to act on an immoral maxim because the will is identified with practical reason, so when we will to perform an immoral act, we implicitly but mistakenly take our underlying policy to be required by reason. By representing our immoral act as rational and reasonable, we are not exercising our powers of reason well, so we are simply making a "choice" that is contrary to reason without "willing" it as such. Our choice is nonetheless free and attributable to us because our will was involved in leading us to take the act to be rational and reasonable. It remains to be seen whether, on this complicated interpretation of Kant, it sufficiently allows for the possibility that one can knowingly and willingly do wrong if the will is practical reason and practical reason is, in part, the moral law.

11. Virtue and Vice

Kant defines virtue as "the moral strength of a *human being's* will in fulfilling his duty" (MM 6:405) and vice as principled immorality (MM 6:390). This definition appears to put Kant's views on virtue at odds with classical views such as Aristotle's in several important respects.

First, Kant's account of virtue presupposes an account of moral duty already in place. Thus, rather than treating admirable character traits as more basic than the notions of right and wrong conduct, Kant takes virtues to be explicable only in terms of a prior account of moral or dutiful behavior. He does not try to make out what shape a good character has and then draw conclusions about how we ought to act on that basis. He sets out the principles of moral conduct based on his philosophical account of rational agency, and then on that basis defines virtue as a kind of strength and resolve to act on those principles despite temptations to the contrary.

Second, virtue is, for Kant, strength of will, and hence does not arise as the result of instilling a "second nature" by a process of habituating or training ourselves to act and feel in particular ways. It is indeed a disposition, but a disposition of one's will, not a disposition of emotions, feelings, desires or any other feature of human nature that might be amenable to habituation. Moreover, the disposition is to overcome obstacles to moral behavior that Kant thought were ineradicable features of human nature. Thus, virtue appears to be much more like what Aristotle would have thought of as a lesser trait, viz. , continence or self-control.

Third, in viewing virtue as a trait grounded in moral principles, and vice as principled transgression of moral law, Kant thought of himself as thoroughly rejecting what he took to be the Aristotelian view that virtue is a mean between two vices. The Aristotelian view, he claimed, assumes that virtue typically differs from vice only in terms of degree rather than in terms of the different principles each involves (MM 6:404, 432). Prodigality and avarice, for instance, do not differ by being too loose or not loose enough with one's means. They differ in that the prodigal person acts on the principle of acquiring means with the sole intention of enjoyment, while the avaricious person acts on the principle of acquiring means with the sole intention of possessing them.

Fourth, in classical views the distinction between moral and non-moral virtues is not particularly significant. A virtue is some sort of excellence of the soul, but one finds classical theorists treating wit and friendliness

alongside courage and justice. Since Kant holds moral virtue to be a trait grounded in moral principle, the boundary between non-moral and moral virtues could not be more sharp. Even so, Kant shows a remarkable interest in non-moral virtues; indeed, much of *Anthropology* is given over to discussing the nature and sources of a variety of character traits, both moral and non-moral.

Fifth, virtue cannot be a trait of divine beings, if there are such, since it is the power to overcome obstacles that would not be present in them. This is not to say that to be virtuous is to be the victor in a constant and permanent war with ineradicable evil impulses or temptations. Morality is “duty” for human beings because it is *possible* (and we recognize that it is possible) for our desires and interests to run counter to its demands. Should all of our desires and interests be trained ever so carefully to comport with what morality actually requires of us, this would not change in the least the fact that morality is still duty for us. For should this come to pass, it would not change the fact that each and every desire and interest *could have* run contrary to the moral law. And it is the fact that they *can* conflict with moral law, not the fact that they actually *do* conflict with it, that makes duty a constraint, and hence is virtue essentially a trait concerned with constraint.

Sixth, virtue, while important, does not hold pride of place in Kant’s system in other respects. For instance, he holds that the lack of virtue is compatible with possessing a good will (G 6: 408). That one acts from duty, even repeatedly and reliably can thus be quite compatible with an absence of the moral strength to overcome contrary interests and desires. Indeed, it may often be no challenge at all to do one’s duty from duty alone. Someone with a good will, who is genuinely committed to duty for its own sake, might simply fail to encounter any significant temptation that would reveal the lack of strength to follow through with that commitment. That said, he also appeared to hold that if an act is to be of genuine moral worth, it must be motivated by the kind of purity of motivation achievable only through a permanent, quasi-religious conversion or “revolution” in the orientation of the will of the sort described in *Religion*. Until one achieves a permanent change in the will’s orientation in this respect, a revolution in which moral righteousness is the nonnegotiable condition of any of one’s pursuits, all of one’s actions that are in accordance with duty are nevertheless morally worthless, no matter what else may be said of them. However, even this revolution in the will must be followed up with a gradual, lifelong strengthening of one’s will to put this revolution into practice. This suggests that Kant’s considered view is that a good will is a will in which this revolution of priorities has been achieved, while a virtuous will is one with the strength to overcome obstacles to its manifestation in practice.

Kant distinguishes between virtue, which is strength of will to do one’s duty from duty, and particular virtues, which are commitments to particular moral ends that we are morally required to adopt. Among the virtues Kant discusses are those of self-respect, honesty, thrift, self-improvement, beneficence, gratitude, sociability, and forgiveness. Kant also distinguishes vice, which is a steadfast commitment to immorality, from particular vices, which involve refusing to adopt specific moral ends or committing to act against those ends. For example, malice, lust, gluttony, greed, laziness, vengefulness, envy, servility, contempt and arrogance are all vices in Kant’s normative ethical theory.

(Interest in Kant’s conception of virtue has rapidly grown in recent years. For further discussion, see Cureton and Hill 2014, forthcoming; Wood 2008; Surprenant 2014; Sherman 1997; O’Neil 1996; Johnson 2008; Hill 2012; Herman 1996; Engstrom 2002; Denis 2006; Cureton forthcoming; Betzler 2008; Baxley 2010).

12. Normative Ethical Theory

The Categorical Imperative, in Kant’s view, is an objective, unconditional and necessary principle of reason that applies to all rational agents in all circumstances. Although Kant gives several examples in the *Groundwork* that illustrate this principle, he goes on to describe in later writings, especially in *The Metaphysics of Morals*, a complicated normative ethical theory for interpreting and applying the CI to human persons in the natural world. His framework includes various levels, distinctions and application procedures. Kant, in particular, describes two subsidiary principles that are supposed to capture different aspects of the CI. The Universal Principle of Right, which governs issues about justice, rights and external acts that can be coercively enforced, holds that “Any action is *right* if it can coexist with everyone’s freedom in accordance with a universal law, or if on its maxim the freedom of choice of each can coexist with everyone’s freedom in accordance with a universal law” (MM 6:230). The Supreme Principle of the Doctrine of Virtue, which governs questions about moral ends, attitudes, and virtue, requires us to “act in accordance with a maxim of

ends that it can be a universal law for everyone to have” (MM 6:395). These principles, in turn, justify more specific duties of right and of ethics and virtue.

In Kant’s framework, duties of right are narrow and perfect because they require or forbid particular acts, while duties of ethics and virtue are wide and imperfect because they allow significant latitude in how we may decide to fulfill them. For example, Kant claims that the duty not to steal the property of another person is narrow and perfect because it precisely defines a kind of act that is forbidden. The duty of beneficence, on the other hand, is characterized as wide and imperfect because it does not specify exactly how much assistance we must provide to others.

Even with a system of moral duties in place, Kant admits that judgment is often required to determine how these duties apply to particular circumstances. Moral laws, Kant says, “must be meticulously observed” but “they cannot, after all, have regard to every little circumstance, and the latter may yield exceptions, which do not always find their exact resolution in the laws” (V 27:574; see also CPR A133/B172; MM 6:411).

13. Teleology or Deontology?

The received view is that Kant’s moral philosophy is a deontological normative theory at least to this extent: it denies that right and wrong are in some way or other functions of goodness or badness. It denies, in other words, the central claim of teleological moral views. For instance, act consequentialism is one sort of teleological theory. It asserts that the right action is that action of all the alternatives available to the agent that has the best overall outcome. Here, the goodness of the outcome determines the rightness of an action. Another sort of teleological theory might focus instead on character traits. “Virtue ethics” asserts that a right action in any given circumstance is that action a virtuous person does or would perform in those circumstances. In this case, it is the goodness of the character of the person who does or would perform it that determines the rightness of an action. In both cases, as it were, the source or ground of rightness is goodness. And Kant’s own views have typically been classified as deontological precisely because they have seemed to reverse this priority and deny just what such theories assert. Rightness, on the standard reading of Kant, is not grounded in the value of outcomes or character.

There are several reasons why readers have thought that Kant denies the teleological thesis. First, he makes a plethora of statements about outcomes and character traits that appear to imply an outright rejection of both forms of teleology. For instance, in *Groundwork* I, he says that he takes himself to have argued that “the objectives we may have in acting, and also our actions’ effects considered as ends and what motivates our volition, can give to actions no unconditional or moral worth...[this] can be found nowhere but in the principle of the will, irrespective of the ends that can be brought about by such action” (G 4: 400). This appears to say that moral rightness is not a function of the value of intended or actual outcomes. Kant subsequently says that a categorical imperative “declares an action to be objectively necessary of itself without reference to any purpose—that is, even without any further end” (G 4:415). A categorical imperative “commands a certain line of conduct directly, without assuming or being conditional on any further goal to be reached by that conduct” (G 4:416). These certainly appear to be the words of someone who rejects the idea that what makes actions right is primarily their relationship to what good may come of those actions, someone who rejects outright the act consequentialist form of teleology. Moreover, Kant begins the *Groundwork* by noting that character traits such as the traditional virtues of courage, resolution, moderation, self-control, or a sympathetic cast of mind possess no unconditional moral worth, (G 4:393–94, 398–99). If the moral rightness of an action is grounded in the value of the character traits of the person who performs or would perform it then it seems Kant thinks that it would be grounded in something of only conditional value. This certainly would not comport well with the virtue ethics form of teleology.

Second, there are deeper theoretical claims and arguments of Kant’s in both the *Groundwork* and in the second *Critique* that appear to be incompatible with any sort of teleological form of ethics. These claims and arguments all stem from Kant’s insistence that morality is grounded in the autonomy of a rational will. For instance, Kant states that “if the will seeks the law that is to determine it anywhere else than in the fitness of its maxims for its own giving of universal law...heteronomy always results” (G 4:441). If the law determining right and wrong is grounded in either the value of outcomes or the value of the character of the agent, it seems it will not be found in the fitness of the action’s maxim to be a universal law laid down by the agent’s own rational will. And Kant’s most complete treatment of value, the second *Critique*’s “On the

Concept of an Object of Pure Practical Reason”, appears to be a relentless attack on any sort of teleological moral theory. “The concept of good and evil” he states, “must not be determined before the moral law (for which, as it would seem, this concept would have to be made the basis) but only (as was done here) after it and by means of it” (CPrR 5:63).

A number of Kant's readers have come to question this received view, however. Perhaps the first philosopher to suggest a teleological reading of Kant was John Stuart Mill. In the first chapter of his *Utilitarianism*, Mill implies that the Universal Law formulation of the Categorical Imperative could only sensibly be interpreted as a test of the consequences of universal adoption of a maxim. Several 20th century theorists have followed Mill's suggestion, most notably, R. M. Hare. Hare argued that moral judgments such as “Stealing is wrong” are in fact universal prescriptions (“No stealing anywhere by anyone!”). And because they are universal, Hare argued, they forbid making exceptions. That in turn requires moral judgments to give each person's wellbeing, including our own, equal weight. And when we give each person's wellbeing equal weight, we are acting to produce the best overall outcome. Thus, in his view, the CI is “simply utilitarianism put into other words” (1993, p. 103). More recently, David Gauthier (1996) has argued that Kant's view that moral principles are justified because they are universalizable is compatible with those principles themselves being consequentialist. Indeed, Gauthier argues that they must be: Respect for the value of humanity entails treating the interests of each as counting for one and one only, and hence for always acting to produce the best overall outcome.

There are also teleological readings of Kant's ethics that are non-consequentialist. Barbara Herman (1993) has urged philosophers to “leave deontology behind” as an understanding of Kant's moral theory on the grounds that the conception of practical reason grounding the Categorical Imperative is itself a conception of value. Herman's idea is that Kant never meant to say that no value grounds moral principles. That, she argues, would imply that there would be no reason to conform to them. Instead, Kant thought the principles of rationality taken together constitute rational agency, and rational agency so constituted itself functions as a value that justifies moral action (1993, 231). Herman's proposal thus has Kant's view grounding the rightness of actions in rational agency, and then in turn offering rational agency itself up as a value. Both Paul Guyer and Allen Wood have offered proposals that differ from Herman's in content, but agree on the general form of teleology that she defends as a reading of Kant. Guyer argues that autonomy itself is the value grounding moral requirements. Moral thinking consists in recognizing the priceless value of a rational agent's autonomous will, something in light of whose value it is necessary for any rational agent to modify his behavior (1998, 22–35). And Wood argues that humanity itself is the grounding value for Kant. While the second *Critique* claims that good things owe their value to being the objects of the choices of rational agents, they could not, in his view, acquire any value at all if the source of that value, rational agency, itself had no value (1999, 130; see also 157–8). Finally, Rae Langton has argued that if Kant's theory is to be thought of as an objectivistic view, we must suppose that the value of humanity and the good will are independent of simply being the objects of our rational choices. If their value thereby becomes the source of the rightness of our actions — say, our actions are right if and because they treat that self-standing value in various ways — then her reading too is teleological.

It is of considerable interest to those who follow Kant to determine which reading — teleological or deontological — was actually Kant's, as well as which view ought to have been his. A powerful argument for the teleological reading is the motivation for Herman's proposal: What rationale can we provide for doing our duty at all if we don't appeal to it's being good to do it? But a powerful argument for the deontological reading is Kant's own apparent insistence that the authority of moral demands must come simply from their being the demands of a rational will, quite apart from the value that will may have (see Schneewind 1996; Johnson 2007, 2008; and Reath 1994). On the latter view, moral demands gain their authority simply because a rational will, insofar as you are rational, must will them. Proponents of this reading are left with the burden of answering Herman's challenge to provide a rationale for having willed such demands, although one response may be that the very question Herman raises does not make sense because it asks, in effect, why it is rational to be rational. On the former view, by contrast, a rationale is at hand: because your will is, insofar as it is rational, good. Proponents of this former reading are, however, then left with the burden of explaining how it could be the autonomy of the will alone that explains the authority of morality.

14. Metaethics

It has seemed to a number of Kant's interpreters that it is important to determine whether Kant's moral philosophy was realist, anti-realist or something else (e.g. a constructivist). This issue is tricky because the terms "realism," "anti-realism" and "constructivism" are terms of art.

One relevant issue is whether Kant's views commit him to the thesis that moral judgments are beliefs, and so apt to be evaluated for their truth or falsity (or are "truth apt").

One might have thought that this question is quite easy to settle. At the basis of morality, Kant argued, is the Categorical Imperative, and imperatives are not truth apt. It makes little sense to ask whether "Leave the gun, take the cannoli." is true. But, in fact, the question is not at all easy. For one thing, moral judgments such as "Lying is wrong" might well be best analyzed according to Kant's views as "The Categorical Imperative commands us not to lie", and this judgment is not an imperative, but a report about what an imperative commands. Thus while at the foundation of morality there would be an imperative which is not truth apt, particular moral judgments themselves would describe what that imperative rules out and so would themselves be truth apt.

Philosophers such as R.M. Hare, however, have taken Kant's view to be that moral judgments are not truth apt. Although on the surface moral judgments can look as if they describe a moral world, they are, as Hare reads Kant, "prescriptions", not "descriptions". This is not, in his view, to say that Kant's ethics portrays moral judgments as lacking objectivity. Objectivity, according to Hare, is to be understood as universality, and the Categorical Imperative prescribes universally.

A second issue that has received considerable attention is whether Kant is a metaethical constructivist or realist.

Constructivism in metaethics is the view that moral truths are, or are determined by, the outcomes of actual or hypothetical procedures of deliberation or choice. Many who interpret Kant as a constructivist claim that his analysis of "duty" and "good will" reveals that if there are moral requirements then the agents who are bound to them have autonomy of the will (Rawls 1980; Korsgaard 1996; O'Neil 1989; Reath 2006; Hill 1989a, 1989b, 2001; Cureton 2013, 2014; Engstrom 2009). Autonomy, in this sense, means that such agents are both authors and subjects of the moral law and, as such, are not bound by any external requirements that may exist outside of our wills. Instead, we are only subject to moral requirements that we impose on ourselves through the operation of our own reason independently of our natural desires and inclinations. The common error of previous ethical theories, including sentimentalism, egoism and rationalism, is that they failed to recognize that morality presupposes that we have autonomy of the will. These theories mistakenly held that our only reasons to be moral derive from hypothetical imperatives about how to achieve given moral ends that exist independently of the activity of reason itself (for a discussion of Kant's more specific objections to previous ethical theories, see Schneewind 2009). On these interpretations, Kant is a skeptic about arbitrary authorities, such as God, natural feelings, intrinsic values or primitive reasons that exist independently of us. Only reason itself has genuine authority over us, so we must exercise our shared powers of reasoned deliberation, thought and judgment, guided by the Categorical Imperative as the most basic internal norm of reason, to construct more specific moral requirements. Kantians in this camp, however, disagree about how this rational procedure should be characterized.

Other commentators interpret Kant as a robust moral realist (Ameriks 2003; Wood 1999; Langton 2007; Kain 2004). According to these philosophers, Kant's theory, properly presented, begins with the claim that rational nature is an objective, agent-neutral and intrinsic value. The moral law then specifies how we should regard and treat agents who have this special status. Autonomy of the will, on this view, is a way of considering moral principles that are grounded in the objective value of rational nature and whose authority is thus independent of the exercise of our wills or rational capacities.

Some interpreters of Kant, most notably Korsgaard (1996), seem to affirm a kind of quietism about metaethics by rejecting many of the assumptions that contemporary metaethical debates rest on. For example, some of these philosophers seem not to want to assert that moral facts and properties just are the outcomes of deliberative procedures. Rather, they seem more eager to reject talk of facts and properties as unnecessary, once a wholly acceptable and defensible procedure is in place for deliberation. That is, the whole framework of facts and properties suggests that there is something we need to moor our moral conceptions to "out there" in reality, when in fact what we only need a route to decision. Once we are more

sensitive to the ethical concerns that really matter to us as rational agents, we will find that many of the questions that animate metaethicists turn out to be non-questions or of only minor importance. Others have raised doubts, however, about whether Kantians can so easily avoid engaging in metaethical debates (Hussain & Shaw 2013).

Bibliography

Primary Sources

Kant's original German and Latin writings can be found in *Königlichen Preußischen Akademie der Wissenschaften* (ed.), 1900–, *Kants gesammelte Schriften*, Berlin: Walter De Gruyter. Most translations include volume and page numbers to this standard Academy edition. Citations in this article do so as well. There are many English translations of Kant's primary ethical writings. The recent Cambridge Edition of the Works of Immanuel Kant provides critical translations of Kant's published works as well as selections from his correspondence and lectures. The following volumes of that series are especially relevant to his moral theory:

Practical Philosophy, translated by Mary Gregor, 1996. Includes: "An Answer to the Question: What is Enlightenment?," *Groundwork of the Metaphysics of Morals*, *Critique of Practical Reason*, and *The Metaphysics of Morals*.

Religion and Rational Theology, translated by Allen Wood and George di Giovanni, 1996. Includes: *Religion within the Boundaries of Mere Reason*

Anthropology, History, and Education, translated by Robert Louden and Guenther Zoeller, 2008. Includes: "Idea for a Universal History with a Cosmopolitan Aim," *Anthropology from a Pragmatic Point of View*, and "Lectures on Pedagogy"

Lectures on Ethics, translated by Peter Heath and J.B. Schneewind, 2001.

Critique of Pure Reason, translated by Paul Guyer and Allen Wood, 1998.

Recent Commentaries on Kant's Ethical Writings

There have been several comprehensive commentaries on the *Groundwork* that have been published recently, some of which also include new English translations.

Allison, Henry, 2011, *Kant's Groundwork for the Metaphysics of Morals: A Commentary*, Oxford: Oxford University Press.

Denis, Lara, 2005, *Groundwork for the Metaphysics of Morals*, Peterborough, Ontario: Broadview Press.

Guyer, Paul, 2007, *Kant's Groundwork for the Metaphysics of Morals: A Reader's Guide*, New York: Continuum.

Hill, Thomas & Zweig, Arnulf, 2003, *Groundwork for the Metaphysics of Morals*, New York: Oxford University Press.

Höffe, Otfried (ed.), 1989, "Grundlegung zur Metaphysik der Sitten": Ein Kooperativer Kommentar, Frankfurt: Vittorio Klostermann.

Schönecker, Dieter, 2015, *Immanuel Kant's Groundwork for the Metaphysics of Morals: A Commentary*, Cambridge, Mass. : Harvard University Press.

Timmermann, Jens, 2007, *Kant's Groundwork of the Metaphysics of Morals: A Commentary*, Cambridge: Cambridge University Press.

Wolff, Robert Paul, 1973, *The Autonomy of Reason: A Commentary on Kant's Groundwork of the Metaphysics of Morals*, New York: Harper & Row.

There are also recent commentaries on the *The Metaphysics of Morals*:

- Andreas Trampota, Andreas, Sensen, Oliver & Timmermann, Jens (eds.), 2011, *Kant's "Tugendlehre": A Commentary*, Berlin: Walter de Gruyter, 343–364.
- Byrd, Sharon & Hruschka, Joachim, 2010 *Kant's Doctrine of Right: A Commentary*, Cambridge: Cambridge University Press.
- Gregor, Mary, 1963, *The Laws of Freedom*, Oxford: Basil Blackwell.

The classic commentary on the *Critique of Practical Reason* is:

Beck, Lewis White, 1960, *A Commentary on Kant's Critique of Practical Reason*, Chicago: University of Chicago Press.

Other Secondary Sources

- Allison, Henry, 1990, *Kant's Theory of Freedom*, New York: Cambridge University Press.
- Ameriks, Karl, 2003, "On Two Non-Realist Interpretations of Kant's Ethics," in his *Interpreting Kant's Critiques*, Oxford: Oxford University Press, 263–282.
- Aune, Bruce, 1979, *Kant's Theory of Morals*, Princeton, NJ: Princeton University Press.
- Baron, Marcia, 2003, "Acting from Duty" in Immanuel Kant, *Groundwork for the Metaphysics of Morals*, Allen Wood (trans.), New Haven: Yale University Press, 98–9.
- , 1995, *Kantian Ethics Almost Without Apology* Ithica: Cornell University Press.
- Baxley, Anne Margaret, 2010, *Kant's Theory of Virtue: The Value of Autocracy*, Cambridge: Cambridge University Press.
- Betzler, Monica, 2008, *Kant's Ethics of Virtue*, New York: Walter de Gruyter.
- Carnois, Bernard, 1987, *The Coherence of Kant's Doctrine of Freedom*, D. Booth (tr.), Chicago: University of Chicago Press.
- Cureton, Adam, 2013, "A Contractualist Reading of Kant's Proof of the Formula of Humanity," *Kantian Review*, 18(3): 363–86.
- forthcoming, "Kant on Cultivating a Good and Stable Will," in Iskra Fileva (ed.), *Perpectives on Character*, Oxford: Oxford University Press.
- & Thomas E. Hill, 2014, "Kant on Virtue and the Virtues," in Nancy Snow (ed.) *Cultivating Virtue*, Oxford: Oxford University Press, 87–110.
- & Thomas E. Hill, forthcoming, "Kant on Virtue: Seeking the Ideal in Human Conditions," in Nancy Snow (ed.), *The Oxford Handbook of Virtue*, Oxford: Oxford University Press.
- , 2014, "Making Room for Rules" *Philosophical Studies*, 172(3): 737–759.
- Darwall, Stephen, 1985, "Kantian Practical Reason Defended," *Ethics*, 96: 89–99.
- , 1977, "Two Kinds of Respect," *Ethics*, 88: 36–49.
- Denis, Lara, 2006, "Kant's Conception of Virtue," in Paul Guyer (ed.), *Cambridge Companion to Kant and Modern Philosophy*, Cambridge: Cambridge University Press, 505–537.
- Engstrom, Stephen, 1992, "The Concept of the Highest Good in Kant's Moral Philosophy," *Philosophy and Phenomenological Research*, 51(4): 747–80.
- , 2009, *The Form of Practical Knowledge*, Cambridge, MA: Harvard University Press.
- , 2002, "The Inner Freedom of Virtue," in Mark Timmons (ed.), *Kant's Metaphysics of Morals: Interpretative Essays*, Oxford: Oxford University Press, 289–315.
- Feldman, Fred, 1978, "Kantian Ethics," in his *Introductory Ethics*, New York: Prentice-Hall, 97–117.
- Foot, Philippa, 1972, "Morality as a System of Hypothetical Imperatives," *The Philosophical Review*, 81(3): 305–316.
- Guyer, Paul, 2000, *Kant on Freedom, Law, and Happiness*, New York: Cambridge University Press.
- , 2005, *Kant's System of Nature and Freedom: Selected Essays*, New York: Cambridge University Press.
- , Guyer, Paul, 2009, "Problems with Freedom: Kant's Argument in Groundwork III and its Subsequent Emendations," in Jens Timmermann (ed.) *Kant's Groundwork of the Metaphysics of Morals: A Critical Guide*, Cambridge: Cambridge University Press, 176–202.
- Herman, Barbara, 1993, *The Practice of Moral Judgment*, Cambridge, MA: Harvard University Press.
- , 1996, "Making Room for Character," in *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, in S. Engstrom and J. Whiting (eds.), New York: Cambridge University Press, 36–60.
- Hill, Thomas E., 2001, "Hypothetical Consent in Kantian Constructivism," *Social Philosophy and Policy*, 18 (2): 300–329. Reprinted in his 2002 *Human Welfare and Moral Worth*, New York: Cambridge University Press, 61–96
- , 1989a, "Kantian Constructivism in Ethics," *Ethics*, 99: 752–70. Reprinted in his 1992 *Dignity and Practical Reason in Kant's Moral Theory*, Ithaca: Cornell University Press, 226–250.
- , 1971, "Kant on Imperfect Duty and Supererogation," *Kant Studien*, 62: 55–76. Reprinted in his 1992 *Dignity and Practical Reason in Kant's Moral Theory*, Ithaca: Cornell University Press, 147–175.
- , 2008, "Kantian Virtue and 'Virtue Ethics'," in Monika Betzler (ed.), *Kant's Ethics of Virtues*, Berlin: Walter De Gruyter, 29–60. Reprinted in his 2012 *Virtue, Rules, and Justice: Kantian Aspirations*, Oxford: Oxford University Press, 129–59.

- , 1973, “The Hypothetical Imperative,” *The Philosophical Review*, 82: 429–50. Reprinted in his 1992 *Dignity and Practical Reason in Kant's Moral Theory*, Ithaca: Cornell University Press, 17–37.
- , 1989b, “The Kantian Conception of Autonomy,” in *The Inner Citadel: Essays on Individual Autonomy*, John Christman (ed.), Oxford: Oxford University Press. Reprinted in his 1992 *Dignity and Practical Reason in Kant's Moral Theory*, Ithaca: Cornell University Press, 76–96.
- , 1972, “The Kingdom of Ends,” *Proceedings of the Third International Kant Congress: Held at the University of Rochester, March 30–April 4, 1970*, Dordrecht: Springer Netherlands, 307–15. Reprinted in his 1992 *Dignity and Practical Reason in Kant's Moral Theory*, Ithaca: Cornell University Press, 58–66.
- Hogan, Desmond, 2009, “Noumenal Affection,” *The Philosophical Review*, 118(4): 501–32.
- Hudson, Hud, 1994, *Kant's Compatibilism*, Ithaca: Cornell University Press.
- Hussain, Nadeem & Shaw, Nishi, 2013, “Meta-ethics and its Discontents: A Case-study of Korsgaard,” in C. Bagnoli (ed.), *Constructivism in Ethics*, New York: Cambridge University Press, 82–107.
- Johnson, Robert N., 1996, “Kant's Conception of Merit,” *Pacific Philosophical Quarterly*, 77: 313–37
- , 2007, “Value and Autonomy in Kantian Ethics,” in *Oxford Studies in Metaethics*, Vol. 2, R. Shafer-Landau (ed.), New York: Oxford University Press, 133–48.
- , 2008, “Was Kant a Virtue Ethicist?,” in *Kant's Ethics of Virtue*, M. Betzler (ed.), Berlin: DeGruyter, 61–76.
- , 2011, *Self-Improvement: An Essay in Kantian Ethics*, New York: Oxford University Press.
- Kain, Patrick, 2004, “Self-legislation in Kant's Moral Philosophy,” *Archiv für Geschichte der Philosophie*, 86(3): 257–306.
- Kohl, Markus, 2016, “Kant on Idealism, Freedom, and Standpoints,” *Archiv für Geschichte der Philosophie*, 98(1): 21–54.
- Korsgaard, Christine, 1996, *Creating the Kingdom of ends*, New York: Cambridge University Press.
- , 1996, *The Sources of Normativity*, O. O'Neill (ed.), New York: Cambridge University Press.
- , 2008, *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, New York: Cambridge University Press.
- Langton, Rae, 2007, “Objective and Unconditioned Value,” *The Philosophical Review*, 116(2): 157–185.
- Louden, Robert, 2000, *Kant's Impure Ethics*, New York: Oxford University Press.
- O'Neill, Onora, 1975, *Acting on Principle*, New York: Columbia University Press.
- , 1989, *Constructions of Reason*, New York: Cambridge University Press.
- , 1996, *Towards Justice and Virtue*, Cambridge: Cambridge University Press.
- Paton, H. J. , 1947, *The Categorical Imperative: A Study in Kant's Moral Philosophy*, London: Hutchinson's University Library.
- Rawls, John, 1980, “Kantian Constructivism in Moral Theory,” *Journal of Philosophy*, 77: 515–72. Reprinted in Rawls, John, 1999, *Collected Papers*, Cambridge, Mass. : Harvard University Press, 303–358.
- , 1989, “Themes in Kant's Moral Philosophy,” in *Kant's Transcendental Deductions*, E. Förster (ed.), Stanford: Stanford University Press, pp. 81–113.
- , 2000, *Lectures in the History of Moral Philosophy*, Cambridge, MA: Harvard University Press.
- Reath, Andrews, 2006, *Agency and Autonomy in Kant's Moral Theory*, Oxford: Clarendon Press.
- , 2015, “Did Kant Hold that Rational Volition is Sub Ratione Boni?,” in Mark Timmons & Robert Johnson (eds.), *Reason, Value, and Respect: Kantian Themes from the Philosophy of Thomas E Hill, Jr.*, New York: Oxford University Press, 232–55.
- , 1994, “Legislating the Moral Law,” *Nöus*, 28(4): 435–464.
- Rippon, Simon, 2014, “Were Kant's Hypothetical Imperatives Wide-Scope Oughts?,” *Australasian Journal of Philosophy*, 92(4): 783–788.
- Ross, W. D., 1954, *Kant's Ethical Theory*, Oxford: Clarendon Press.
- Schmucker, Josef, 1961, *Die Ursprünge der Ethik Kants*, Meisenheim am Glan: Anton Hain.
- Schneewind, J. B. , 2009, “Kant Against the ‘spurious principles of morality’,” in J. Timmermann (ed.), Cambridge: Cambridge University Press.
- , 1996, “Kant and Stoic Ethics,” in S. Engstrom and J. Whiting (eds.) *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, Cambridge: Cambridge University Press, 285–301.
- Schroeder, Mark, 2005, “The Hypothetical Imperative?,” *Australasian Journal of Philosophy*, 83(3): 357–372.
- Sensen, Oliver, 2013, “Kant's Constructivism” in C. Bagnoli (ed.), *Constructivism in Ethics*, New York: Cambridge University Press, 63–81.


- , 2011, “Kant on Duties Toward Others From Respect (TL 37–44),” in Andreas Trampota, Oliver Sensen and Jens Timmermann (eds.), *Kant's “Tugendlehre”: A Commentary*, Berlin: Walter de Gruyter, 343–364.
- Sherman, Nancy, 1998, *Making a Necessity of Virtue*, New York: Cambridge University Press.
- Sullivan, Roger J., 1989, *Immanuel Kant's Moral Theory*, New York: Cambridge University Press.
- Surprenant, Chris, 2014, *Kant and the Cultivation of Virtue*, New York: Routledge.
- Timmons, Mark (ed.), 2002, *Kant's Metaphysics of Morals: Interpretative Essays*, New York: Oxford University Press.
- Wolff, Robert Paul (ed.), 1967, *Kant: A Collection of Critical Essays*, Garden City, New York: Doubleday.
- (ed.), 1969, *Kant's Foundations of the Metaphysics of Morals: Text and Commentary*, Indianapolis, IN: Bobbs-Merrill.
- Wood, Allen, 1970, *Kant's Moral Religion*, Ithaca: Cornell University Press.
- , 1999, *Kant's Ethical Thought*, New York: Cambridge University Press.
- , 2007, *Kantian Ethics*, New York: Cambridge University Press.

Academic Tools

 [How to cite this entry.](#)

 [Preview the PDF version of this entry](#) at the [Friends of the SEP Society](#).

 [Look up this entry topic](#) at the [Indiana Philosophy Ontology Project](#) (InPhO).

 [Enhanced bibliography for this entry](#) at [PhilPapers](#), with links to its database.

Other Internet Resources

- [Kant and Kantian Ethics](#), at Ethics Updates, maintained by Lawrence Hinman
- [Bibliography of secondary literature on Kant's Ethics \(PDF\)](#) maintained by Jorg Schroth
- [Kant on the Web](#), maintained by Stephen Palmquist.
- [Kant Papers](#), articles, links, news and reviews about Kantian philosophy.

Related Entries

[autonomy: personal](#) | [character, moral](#) | [consequentialism](#) | [consequentialism: rule](#) | [constructivism: in metaethics](#) | [ethics: deontological](#) | [ethics: virtue](#) | [Kant, Immanuel](#) | [Kant, Immanuel: account of reason](#) | [Kant, Immanuel: aesthetics and teleology](#) | [Kant, Immanuel: and Hume on morality](#) | [Kant, Immanuel: philosophical development](#) | [Kant, Immanuel: philosophy of religion](#) | [Kant, Immanuel: social and political philosophy](#) | [Kant: transcendental idealism](#) | [morality, definition of](#) | [practical reason](#) | [respect](#) | [rights](#)

Copyright © 2016 by

Robert Johnson <JohnsonRN@missouri.edu>

Adam Cureton <adamcureton@utk.edu>

Open access to the Encyclopedia has been made possible by a [world-wide funding initiative](#). See the list of [contributing institutions](#). If your institution is not on the list, please consider asking your librarians to contribute.

Stanford | Center for the Study of
Language and Information

The Stanford Encyclopedia of Philosophy is [copyright © 2016](#) by [The Metaphysics Research Lab](#), Center for the Study of Language and Information (CSLI), Stanford University

Library of Congress Catalog Data: ISSN 1095-5054