

KANT'S MORAL PHILOSOPHY

by Robert Johnson, *The Stanford Encyclopedia of Philosophy*

Kant argued that moral requirements are based on a standard of rationality he dubbed the “Categorical Imperative” (CI). Immorality thus involves a violation of the CI and is thereby irrational. Other philosophers, such as Locke and Hobbes, had also argued that moral requirements are based on standards of rationality. However, these standards were either desire-based instrumental principles of rationality or based on *sui generis* rational intuitions. Kant agreed with many of his predecessors that an analysis of rationality will reveal only a requirement to conform to instrumental principles. Yet he 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 this 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 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* and *Religion within the Boundaries of Mere Reason*. I will 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. I myself still find the standard approach most illuminating, though I will highlight important positions from the later works where needed.

I. Aims and Methods of Moral Philosophy

The first fundamental aim of moral philosophy, and so also of the *Groundwork*, is to “seek out” the foundational principle of a metaphysics of morals. Kant pursues this project through the first two chapters of the *Groundwork*. He proceeds by analyzing and elucidating commonsense ideas about morality. The point of this project was 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 here are supposed to be those any normal, sane, adult human being would accept, although nowadays Kant would be widely regarded as overly optimistic regarding the depth and extent of moral agreement. He is perhaps 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 take himself to be addressing the genuine moral skeptics who often populate the works of moral philosophers, that is, people who need a reason to act morally and whose moral behavior hinges on a rational proof that philosophers might try to give. Thus, when, in the third and final chapter of the *Groundwork*, Kant takes up his second fundamental aim, to “establish” this fundamental moral principle and show that it is a principle

of rationality, his conclusion is not meant to answer a skeptical challenge. Kant rests this second project on the position that we — or at least creatures with rational wills — possess autonomy. His argument in this regard does often appear to try to reach out to a metaphysical fact about our wills. And this may lead readers to the conclusion that he is, after all, trying to justify moral requirements by appealing to a fact — our autonomy — that even a radical moral skeptic would have to recognize. However, the most important elements of his argument to establish the fundamental principle of morality rest on a claim that will be less impressive to a skeptic, that the autonomy of our wills is a presupposition of any practical point of view recognizable as such by us.

Although these are the two fundamental aims of moral philosophy, they are not, in Kant's view, the only aims. Moral philosophy addresses the question, What ought I to do?, and an answer to that question requires much more than delivering the fundamental principle of morality. We also need some account, presumably based on this principle, of the nature and extent of our various ethical obligations. To this end, Kant employs his findings from the *Groundwork* in *The Metaphysics of Morals*, offering a categorization of our basic ethical obligations to ourselves and others. Moral philosophy should also characterize and explain the demands morality makes on human psychology and forms of human social interaction. These topics, among others, are addressed in central chapters of the second *Critique* and again in the *Metaphysics of Morals*, but are perhaps given their most 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 ethical 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, virtue does not insure wellbeing and may even conflict with it. Further, 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, 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 Divine Providence) 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 must be addressed *a priori*, that is, without drawing on observations of human beings and their behavior. Once we “seek out and establish” the fundamental principle of morality *a priori*, then we may consult facts drawn from experience in order to determine how best to apply this principle to human beings and generate particular conclusions about how we ought to act. 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 *Groundwork*, for instance, makes repeated appeals to empirical facts (that our wills are determined by practical principles, that various motivations are variable in producing right actions, and so on).

Later ethical works rely even more heavily on empirical generalizations. Kant did not take himself to be employing these assumptions in seeking out and establishing the fundamental moral principle, only in applying it to human beings. Nevertheless, it is not always easy to tell whether Kant's arguments gain their plausibility only by relying on ideas established by observations of human being and the world they inhabit.

In one sense, it might seem obvious why Kant nevertheless insists on an *a priori* method. A 'metaphysics of morals' would be, more or less, an account of the nature and structure of moral reality — in effect, a categorization of duties and values. Such a project would answer 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. A principle used to provide such categorizations thus appears to be a principle of metaphysics. And metaphysical principles 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 Kant conceives of ethics, as have other philosophers, as initially requiring 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 a 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, the most important project is to show that we, as rational agents, are justified in holding ourselves to the standard enshrined in the CI. If this project must also be carried out *a priori*, it cannot be because it is an analytical project. Conformity to moral requirements is not itself contained in the concept of rational agency, nor is there any contradiction in asserting that a fully rational agent can be immoral. Yet, Kant argued, no experience can tell us whether any will, including our own, conforms to the CI either. "What counts", regarding such conformity, "is not the actions one sees, but their inner principles, which one does not see." (4:407) Indeed, because of the dismal record of human behavior, the more we experience it, the less confidence we have that any will ever conforms to moral norms. Since observation cannot establish the necessary conformity of rational wills to the CI, Kant regards the claim that they do conform as an example of an *a priori synthetic* claim, an *a priori* claim that is not analytic or conceptual, yet whose justification cannot rely on observation. This is the second reason Kant held 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 the third reason for an *a priori* method, and it appears to have been of great importance to Kant: Moral require-

ments present themselves as being *absolutely necessary*. But an *a posteriori* method seems ill-suited to discovering and establishing what we *must* do; surely it will 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 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 absolute 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 absolutely 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 absolute 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 in fact 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 he returns throughout his works. The basic idea 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 only makes decisions that she holds to be morally worthy, taking 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 believes we value it without limitation or qualification. By this, I 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 determination to give moral considerations decisive weight is worth honoring, *but only under such and such circumstances*.

Second, as a consequence, possessing and maintaining one's moral goodness 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 a commitment to honor 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 it 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 or overall welfare.

In Kant's terms, a good will is a will whose decisions are wholly determined by moral demands or as he refers to this, by the Moral Law. Human beings view this Law as a constraint on their desires, and hence a 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. A holy will would be entirely free from desires that might operate independently of morality. 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, or the thought that we have moral duties.

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 content of one's will. 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 under any circumstances. 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 the motivational structure of the agent be arranged so as to give considerations of duty priority over all other interests. It does not require or even recommend a rule-bound character devoid of the warmth of human emotion.

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 lawfulness. What naturally comes to mind is this: Duties are created by rules or laws of some sort. For instance, the bylaws of a club lay down duties for its officers. City and state laws establish the duties of citizens. 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 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 never 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 in acting from duty that we are not at all motivated by a prospective outcome or some other extrinsic feature of our conduct. We are motivated by the mere conformity of our will to law as such.

What, then, is the difference between being motivated by a sense of duty in the ordinary sense, and being motivated by duty in Kant's sense? It is, presumably, this: Motivation by duty is motivation by our respect for whatever law it is that makes our action a duty. But we can rationally 'opt out' of our membership in the city, state, club or any other social arrangement and its laws — for instance, by quitting the club or expatriating. Those laws only apply to us given we don't rationally decide to opt out, given the opportunity. Our respect for the laws guiding us is qualified, in the sense that the thought that the law gives us a duty is compelling only if there is no law we respect more that conflicts with it: My respect for the laws of my club guides my action only insofar as those laws do not require me to violate city ordinances. But my respect for city ordinance guides me only insofar as they do not require me to violate federal law. And so on.

Eventually, however, we come to laws that apply to us simply as members of the 'club' of rational agents, so to speak, as beings who are capable of guiding their own behavior on the basis of directives, principles and laws of rationality. We cannot choose to lay aside our 'membership' in the category of such beings, or at least it is unclear what the status of such a choice would be. So, suppose that there is some law prescribing what any rational agent must do. Then we have an idea of a duty that we cannot rationally opt out of. When we do something because it is our moral duty, Kant argued, we are motivated by the thought that, insofar as we are rational beings, we must act only as this fundamental law of (practical) reason's prescribes, a law that would of how any rational being in our circumstances should act. Whatever else such a law might be, it is, in virtue of being a principle of reason, universally valid. My respect for such a law is thus not qualified: my respect for the laws of my club, city, constitution or religion guides me in practical affairs only insofar as they do not require me to violate laws laid down by my own practical reason, but my respect for the deliverances of my own reason does not depend on whether it requires me to violate the former sorts of laws. In this case, it is respect for (rational) lawfulness as such guides me.

The forgoing line of argument reveals a distinctive aspect of Kant's approach: his 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. Since they retain their reason-giving force under any circumstance, they have universal validity. So, whatever else may be said of 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' (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 at the basis of all of our moral duties is a *categorical imperative*. It is an *imperative* because it is a command (e.g., "Leave the gun. Take the cannoli.") More precisely, it commands us to exercise our *wills* in a particular way, not to perform some action or other. 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. Of course, other imperatives have a similar non-conditional form. For instance, 'Answer an invitation in the third person in the third person' is an imperative of etiquette, and it is not conditional. (Foot, 1972, p. 308) It does not apply to you only on the condition that you have some end that is served by being polite. But this imperative is not *categorical* in Kant's sense, since it does not apply to us simply because we are rational enough to understand and act on it, or simply because we possess a rational will. Imperatives of etiquette apply to us simply because prevailing customs single us out as appropriate objects of appraisal by standards of politeness, whether we accept those standards or not.

There are 'oughts' other than our moral duties however, but these oughts are distinguished from the moral ought in being based on a quite different kind of principle, a principle that is a *hypothetical imperative*. 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 does not posit any end that you will, but consists 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 or wanting it; it requires the exercise of practical reason and focusing oneself on the pursuit of that end. That is, an imperative that applies to us because we desire some end is not 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 an end that we *must* will. The distinction between ends that we might or might not will and those, if any, we *must* will, 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 again.” 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 *must* will.

As it turns out, the only (non-moral) end that we *must* will in Kant's view (by ‘natural necessity’ he says) is our own happiness. Any imperative that applied to us because we will our own happiness would thus be an *assertoric* imperative. As it turns out, however, rationality 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. 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, applying to the actions themselves only derivatively. 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 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 on our willing.

Kant describes the will as operating on the basis of subjective volitional principles he calls ‘maxims’. Hence, morality and other rational requirements are demands that apply to the maxims that motivate our actions. The form of subjective principles of willing 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. Since this is a principle stating only what some *agent* wills, it is *subjective*. (A principle for *any* rational will would be 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 must be rational.

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 (1989, 1999), among others, take this formulation

in effect to summarize a decision procedure for moral reasoning, and I will follow them: 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 it. (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 (1992), 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 enjoin the pursuit of an end, come in the form '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. 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, we recognize 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, 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 an imperfect duty toward others. Again, Kant's interpreters differ over exactly how to reconstruct the derivation of these duties. I 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 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 it gets 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 it will get what they want, I am conceiving of a world in which no practice of giving one's word could ever arise. So I am conceiving of a world in which 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, 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 my maxim exists together with itself as a universal law. Since it is inconceivable that these two things should 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 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* talent 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 immoral. In order to show that this maxim is categorically forbidden, I believe we must 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. (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. (4:423) He also appears to rely on this claim in each of his examples. Each maxim appears to have happiness as its aim. One explanation for this is that, since each person necessarily wills 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” rather than developing one's 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 capacities. We cannot do so, because our own happiness is the very end contained in the maxim of giving oneself 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 someone sometime develop talents. So since we cannot will as a universal law of nature that no one ever develop any talents — given 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 do this all the time. 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 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 distinc-

tively human, and these include capacities to engage in self-directed rational behavior and to adopt and pursue our own ends, and any other 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 when 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 produce. An end in this sense guides my actions in that once I will to produce something, I then deliberate about means of producing it. 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 will act in ways that will bring about that end.

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 self-preservation is a *subjective* end, while Humanity is an *objective* end. Self-preservation is subjective in that it is not an end that every rational being must have. We do place more importance on it than most of our other positive ends. Because self-preservation is more important to me than excitement, I am not a base-jumper, and so self-preservation puts a limit on my behavior. But I could make self-preservation less important if I wish, and perhaps put excitement in its place so that it, and not self-preservation, limits pursuit of my other ends. Humanity is an *objective* end, because it is an end that every rational being must have insofar as she is rational. Hence, it limits what I am *morally permitted* to do when I pursue my positive *and* subjective negative 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's 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 take

account of their ends in my own plans. In so doing, I further the Humanity in others, by helping further the projects and ends the adoption and pursuit of which constitutes that Humanity. It is this sense of Humanity as an end-in-itself on which Kant's arguments for imperfect duties rely.

Finally, Kant's 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 consistent with 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, cases of respecting a person because of who or what that person is, 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.

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*." (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 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 to bring 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, 1992). Many see in 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”. (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” (4:439). It combines the others in that (i) it requires that we conform our actions to the maxims of a legislator of laws (ii) that this lawgiver 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”. (4:435). He also says that one formula “follows from” another (4:431), and that the concept foundational to one formula “leads to a closely connected” concept at the basis of another formula (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. Perhaps Kant thought this, but it is not very plausible: That I should always treat Humanity as an end in itself, for instance, does not seem to mean the same thing as that I should act only on maxims that are consistent with themselves as universal laws of nature.

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. (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*. (4:429n) One might take this as 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 (I 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. 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 is their differences in meaning, 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 position that rational human wills are autonomous. Kant saw this as the key to understanding and justifying the authority 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* influences 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.

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 reasons. This is, firstly, the concept of a will that does not operate through the influence of factors outside of this responsiveness to reasons. For a will to be free is thus for it to be physically and psychologically unforced in its operation. Hence, choices made 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 (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 casual 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 Xs to Ys . 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 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, 1992), 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 existing 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 uses this to argue that our wills are autonomous. Hence, while in the *Groundwork* Kant argues from a dubious argument for our autonomy to our being 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 and hypothetical. The conclusions are thus fully compatible with morality being, as he puts it, a “mere phantom of the brain”. (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’. 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*. (5:57-67)

II. Virtue and Vice

Kant defines virtue as “the moral strength of a *human being's* will in fulfilling his duty” (6:405) and vice as principled immorality. (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 the trait of acting according to these principles.

Second, virtue is for Kant a 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 differs from vice only in terms of degree rather than in terms of the different principles each involves. (6:404, 432) But 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 acts on the principle of acquiring means with the sole intention of enjoyment, while the avaricious act 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 along side 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. 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 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. (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*. Kant here describes the natural human condition as one in which no decisive priority is given to the demands of morality over happiness. 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.

Bibliography

- Allison, Henry, 1990, *Kant's Theory of Freedom*. New York: Cambridge U. P.
- Aune, Bruce, 1979, *Kant's Theory of Morals*. Princeton, NJ: Princeton U. P.
- Baron, Marcia, 1995, *Kantian Ethics Almost Without Apology* Cornell: Cornell U.P.
- Beck, Lewis White, 1960, *A Commentary on Kant's "Critique of Practical Reason"* Chicago: University of Chicago Press.
- Carnois, Bernard, 1987, *The Coherence of Kant's Doctrine of Freedom*. D. Booth, tr. Chicago: University of Chicago Press.
- Darwall, Stephen, 1985, “Kantian Practical Reason Defended”. *Ethics* 96; pp. 89-99
- -----, 1977, “Two Kinds of Respect,” *Ethics* 88: 36-49.
- Engstrom, Stephen. 1992, “The Concept of the Highest Good in Kant's Moral Philosophy” *Philosophy and Phenomenological Research* 51, no. 4; pp. 747-80
- Feldman, Fred. 1978, “Kantian Ethics” from his *Introductory Ethics* New York: Prentice-Hall; pp. 97-117.
- Gregor, Mary, 1963, *The Laws of Freedom*. Oxford: Basil Blackwell.
- Foot, Philippa, 1972, “Morality as a System of Hypothetical Imperatives,” *The Philosophical Review* , Vol. 81, No. 3. (Jul., 1972), pp. 305-316.
- Guyer, Paul, 2000, *Kant on Freedom, Law, and Happiness*. New York: Cambridge U. P.
- -----, ed., 1998, *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*. Lanham, MD: Rowman and Littlefield.
- Herman, Barbara, 1993, *The Practice of Moral Judgment*. Cambridge, MA: Harvard U. P.
- Hill, Thomas E., 1992, *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca: Cornell U. P.
- -----, 2002, *Human Welfare and Moral Worth*. New York: Cambridge U. P.
- -----, & Zweig, A., eds., Zweig, A., tr., “Editors' Introduction” and “Analysis of Arguments” in *Kant: Groundwork for the Metaphysics of Morals*. New York: Oxford U. P., 2003; pp. 19-177.
- Höffe, Otfried, ed., 1989, “*Grundlegung zur Metaphysik der Sitten*”: *Ein Kooperativer Kommentar*. Frankfurt: Vittorio Klostermann.

- Johnson, Robert N., 1996, "Kant's Conception of Merit," *Pacific Philosophical Quarterly* 77; pp. 313-37
- Korsgaard, Christine, 1996, *Creating the Kingdom of ends*. New York: Cambridge U. P.
- -----, 1996, *The Sources of Normativity*. O. O'Neill, ed., New York: Cambridge U. P.
- Louden, Robert, 2000, *Kant's Impure Ethics*. New York: Oxford U. Press.
- O'Neill, Onora, 1975, *Acting on Principle*. New York: Columbia U. Press.
- -----, 1989, *Constructions of Reason*. New York: Cambridge U. Press.
- Paton, H. J., 1947, *The Categorical Imperative: A Study in Kant's Moral Philosophy*. London: Hutchinson's University Library.
- Rawls, John, 1989, "Themes in Kant's Moral Philosophy," in *Kant's Transcendental Deductions*, E. Förster, ed., pp. 81-113. Stanford: Stanford U. P.
- -----, 2000, *Lectures in the History of Ethics*. Cambridge, MA: Harvard U. P.
- Reath, Andrews, 1994, "Legislating the Moral Law" *Nôus* 28, No. 3, Dec.
- Ross, Sir David, 1954, *Kant's Ethical Theory*. Oxford: Clarendon Press.
- Schmucker, Josef, 1961, *Die Ursprünge der Ethik Kants*. Meisenheim am Glan: Anton Hain.
- Sherman, Nancy, 1998, *Making a Necessity of Virtue*. New York: Cambridge U. P.
- Sullivan, Roger J., 1989, *Immanuel Kant's Moral Theory*. New York: Cambridge U. P.
- Timmons, Mark, ed., 2002, *Kant's Metaphysics of Morals: Interpretative Essays*. New York: Oxford U. P.
- Wolff, Robert Paul, 1973, *The Autonomy of Reason*. New York: Harper.
- -----, 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, NY: Cornell U. P.
- -----, 1999, *Kant's Ethical Thought*. New York: Cambridge U. P.