

The Formula of Humanity as an End in Itself

The “humanity formulation” of the Categorical Imperative demands that every person must

Act in such a way that you treat humanity, whether in your own person or in any other person, always at the same time as an end, never merely as a means.
(4:429)¹

Of the different formulations of the Categorical Imperative, or different ways of stating the fundamental principle of morality, the humanity formulation is probably the most intuitively compelling. Yet, despite its intuitive appeal, even the most basic elements of the humanity formulation – what exactly must be treated as an end in itself, what is involved in treating something as an end in itself, and why we should accept this basic moral requirement – are surprisingly unclear and even controversial. My aim in this essay is to offer a consistent account of these basic issues, while acknowledging alternative interpretations and preserving the intuitive force of Kant’s principle.

Although there is significant disagreement about the details of the humanity formulation, there is at least rough consensus among commentators about some starting points.

One generally accepted interpretive idea is that Kant is not saying that exactly all and only members of the human species must be treated as ends in themselves. His use of the term “humanity” (in German, “*die Menschheit*”) is potentially misleading, since he also consistently says that “rational beings” are ends in themselves, in virtue of their “rational nature.” Rationality is the key feature that distinguishes typical humans from all the other beings that we know of, so it is not too great a slip to say that human beings are ends in themselves. But Kant also thinks that God possesses a rational nature, more perfectly rational than our own, and he even thinks it

could well be that there are races of rational beings living on other planets (7:332). So Kant does not mean to say that only members of our biological species can be rational or must be treated as ends in themselves. And if rational nature is what makes a being an end in herself, then being biologically human is not a sufficient condition for being an end in oneself either. Some humans – patients in permanent vegetative states, for example – lack even the most minimal sort of rationality. So there is a consensus that Kant means that rational beings are ends in themselves, not that all and only members of the species *homo sapiens* are ends in themselves.

There also is some agreement about what it means to call something an end in itself. For Kant, an end (*der Zweck*) is a reason or purpose for action. Most ends are contingent ends, based on desires or feelings that may vary from person to person. For instance, learning to speak Chinese may be an end for you but not for me, because we have different plans and desires. No person is rationally required to adopt any contingent end – there is no reason for me to adopt the end of learning Chinese if I have no desire to and if doing so does not serve any further purpose I have. But besides contingent ends, there is another kind of end that necessarily provides a compelling reason for every rational agent to act in certain ways. This is an end in itself. There are some requirements or principles built into any rational deliberation about what to do – some considerations that we can not rationally ignore -- and Kant expresses this by saying that there is a necessary end that must be given weight in any agent's deliberations. This end in itself is rational nature, in oneself and in others. Because of rational nature's special status, we can not regard it as having significance only because of what it gets us. So, we must not treat it as merely a means to the satisfaction of our own desires.

This gives an abstract idea of what the phrase “end in itself” means, but Kant of course goes further, and specifies some of the ways in which we are required to treat rational nature

because of its being an end in itself. The unique importance of rational nature, according to Kant, leads to specific duties to develop our own powers of rationality, to refrain from destroying ourselves, to give some weight to others' choices and concerns, and to treat others with respect. How exactly such duties follow from the idea of rational nature's special status is often unclear in Kant's texts, and requires some speculative reconstruction of his arguments.

So, some basic points regarding the humanity formulation are clear enough. The humanity formulation requires that rational nature, and every rational being, must be given a special weight in everyone's deliberations about how to act. And Kant enumerates some particular duties that are implied by this special status. But each of these points needs further explanation. And even after Kant's view is explained, the question remains of why we ought to accept any of it. I turn now to exploring these points in more detail.

What Should We Treat as an End in Itself?

It is widely, and I think rightly, agreed that Kant thinks the end in itself is rational nature. But the phrase "rational nature" is ambiguous.

In part because of the influence of economic models of rationality, rationality nowadays is associated most naturally with instrumental reason, or with calculating the most effective means to satisfy one's own desires. But Kant does not use the words "reason" or "rationality" (the same German word, *die Vernunft*, is translated into either English word) or the adjective "rational" (*vernunftige*) to be associated mainly with instrumental reason. So the characteristic that makes someone an end in herself is not the ability to seek effective means to her ends. Neither does Kant mean to contrast "rational" with "emotional," so the humanity formulation is

not saying that beings who suppress their emotions have a special moral status that others lack.

Kant's conception of reason or rationality should not be taken to have the same sense or connotation as these current versions of "rationality," since Kant's own account of the power of reason is perhaps the most distinctive and fundamental element of his overall philosophical system. Kant divides the power of reason into theoretical and practical aspects. In either aspect, reason is a very active faculty. In its theoretical use, reason supplies principles that guide the understanding in its task of organizing our sense impressions into coherent and understandable patterns. Theoretical reason does not just respond to passively received information, but instead spontaneously provides principles that make coherent perception and empirical scientific investigation possible. Similarly, in its practical use, reason does more than seek the best means to whatever contingent desires we passively find ourselves possessing. Practical reasoning has to do with the exercise of our will. One aspect of our will is *Willkür*, or the power to make choices about which ends we will adopt. Another aspect is *Wille*, which presents or "legislates" categorical moral principles to an agent. Every competent adult human, and any other rational being that may exist, has both *Willkür* and *Wille*, on Kant's picture. A perfectly rational agent would always exercise her power of choice, or *Willkür*, in ways that are consistent with the moral demands given by *Wille*, because her own power of reason presents these moral principles as unconditional reasons for action. But of course, actual people are imperfectly rational, and so they sometimes choose to perform actions that are contrary to self-given moral principles.

Kant's account of reason and the will presents many possible candidates for the end in itself. Perhaps it is theoretical reason that is the end in itself, or perhaps it is practical reason. Or maybe some particular aspect of practical reason, such as a being's *Willkür*, qualifies her as an end in herself. Or maybe someone is an end in herself in virtue of the power to legislate moral

principles. Or, to adopt a more demanding requirement, maybe someone is an end in herself only if she not only possesses *Willkür* and *Wille*, but also is sufficiently rational to recognize and acknowledge the unconditional force of moral principles, and so regulates her particular choices in accordance with moral principles. In fact, there is no clear consensus about which of these possible readings of “rational nature” as an end in itself is correct, and almost every version has received some support in recent decades. Although there has been surprisingly little attention paid to resolving the divergence in the readings of the “humanity” or “rational nature” that must be treated as an end in itself, the issue seems crucial to a proper understanding of the humanity formulation.

Some commentators have proposed that the end in itself is the power to set ends (*Willkür*), or the power to set ends plus the power to organize one’s ends into a consistent package (Korsgaard, 1996, p. 17, p. 110, p. 346, and Wood, 1999, p. 118-120). The importance of the power of choice is a familiar idea to current moral philosophers, steeped in political liberalism’s emphasis on the foundational importance of choice, but given Kant’s emphasis on the importance of choosing to obey reason’s moral demands, it would be odd if the power of choice *simpliciter* were the end in itself that must be given special consideration. The power of choice is only one aspect of practical reason, and someone can possess the power of choice and yet choose to act contrary to the rationally compelling principles legislated by *Wille*, another aspect of her own power of reason. In light of this, identifying the power of choice as “rational nature” would seem peculiar.

For similar reasons, it would be even stranger if Kant took just the use of theoretical reason as the main distinguishing feature of the end in itself. While the exercise of theoretical reason does culminate in impressive theories of physical science and in some spectacular

technical accomplishments, it leaves out an aspect of reason that Kant consistently emphasizes, namely the aspect related to moral reasoning. In *Groundwork*, Kant says that “Morality is the only condition under which a rational being can be an end in himself... Therefore morality, and humanity so far as it is capable of morality, is the only thing which has dignity” (4:435). In the later work *The Metaphysics of Morals* (which develops and applies the basic moral principles of *Groundwork*), Kant says the respect we should show another person is “respect for man as a moral being” (6:464), and that “morality is an end in itself” (6:422-23, also 6:436). In *Critique of Practical Reason*, Kant says that “the human being (and with him every rational being) is an end in itself” because a human being “is the subject of the moral law and so of that which is holy in itself, on account of which and in agreement with which alone anything can be called holy” (5:131-32). In the *Critique of Judgment*, Kant says that “...it is only as a moral being that man can be the final purpose [end, or *Zweck*] of creation” (5:437, 5:443). The aspect of rational nature that marks a being as an end in herself, then, is the aspect related to legislating and acting on moral principles.

But this still is not a precise answer. One feature of rational nature which makes moral reasoning possible is the *Wille*'s activity of legislating moral principles. Universal moral principles are built into the activity of deliberating about what to do, so every agent must take her own power of practical reason to be presenting her with unavoidable moral requirements.² Perhaps the power of legislating unconditional moral demands is the aspect of reason that marks a rational being as an end in herself.

But some commentators have taken the “capacity for morality” to be the distinguishing feature of beings who are ends in themselves, and this capacity involves more than just legislating moral principles. For a perfect or “holy will” like God's, the requirements of reason

might be sufficient to lead to moral actions, but for beings like us, who are influenced by sensual desires and emotions, there must also be some feeling that accompanies the choice to act on the moral law. The feeling of *Achtung*, translated as “respect” or “reverence,” is the key feeling that is needed. Kant distinguishes *Achtung* from the typical sort of human feeling, in that it is “not of empirical origin,” instead being “produced by an intellectual ground,” namely the recognition of moral principles’ unconditional power to command. “Moral law strikes down self-conceit” (5:73) by showing that there is something more important than our own inclinations, and “what in our own judgment infringes upon our self conceit humiliates” (5:74). But it also produces a positive feeling of *Achtung* for the moral law itself, because the “relative weightiness of the law” is made apparent by its “removal of the counterweight” (5:76) of immoral desires. So although the moral law can produce a “feeling of displeasure” by “the lowering of pretensions to self-esteem,” (5:78-79) it also is “an elevation of the moral” and as such there is “so little displeasure in it that, once one has laid self-conceit aside ... one can in turn never get enough of contemplating the majesty of this law” (5:77). *Achtung*, then, is a moral feeling, a positive feeling of respect for moral principles which is inspired by the objective normative force of such principles. For beings like us, who are rational yet subject to sensual influences, the capacity for morality is most plausibly taken to be the combination of *Wille*, *Willkür*, and the moral feeling of *Achtung*, which can counteract feelings and desires that may lead us away from morality’s demands. Some commentators have defended the view that this capacity for morality may be what qualifies a being as an end in herself.³

But to merely possess the capacity for morality is not to be rational in the fullest sense possible for human beings. Moral principles are commanded by *Wille*, an aspect of our own power of practical reason, as sufficient reasons for action. So someone who acts contrary to

moral principles is acting irrationally, on Kant's picture. It may well be that by saying that rational nature is an end in itself, Kant means to say that a being is an end in herself only if she is committed to acting on the moral principles demanded by her own power of reason. This reading of the end in itself is certainly not conventional among Kant commentators, mostly because it appears to have unpalatably moralistic consequences -- it appears to demand judgments of others' moral character, and differential treatment based on those judgments. But I argue elsewhere that it does not have the repugnant consequences that one might think, and that it is the most textually justified reading of the humanity formulation (Dean, 1996, and Dean, 2006).

We can provisionally conclude that the end in itself is rational nature, where rational nature includes some aspect of moral reason – either the power to legislate moral principles, the overall capacity to act on moral principles, or the commitment to actually act on them. Which of these readings is most justified depends largely on which one allows the most compelling and plausible reconstructions of Kant's basic arguments related to the humanity formulation.

Value and Ends

Before turning to the argument for and applications of the humanity formulation, some explanation is needed of Kant's conception of value.

For Kant, the choices of rational agents are conceptually prior to the value of the objects chosen, so the valuable is that which is an "object of practical reason" (5:57). This no doubt sounds odd to many readers. The influence of consequentialism in moral philosophy, and the influence of rational choice theory in many disciplines, makes it seem natural to take rational choice as a matter of reacting properly to preexisting value, perhaps even by maximizing that

which has value. But Kant explicitly denies this. Value does not present itself to an agent as a result of any preexisting external or internal state. In his most extended discussion of the concept of value, in *Critique of Practical Reason* 57-64, Kant first argues that the fact that someone wants something, or prefers some state of affairs to come about, is not sufficient to show that the state of affairs is of value. Preferences or desires by themselves are not sufficient conditions for value, they are only of value when they are consistent with one's overall happiness, because a person's "reason certainly has a commission from the side of his sensibility which it cannot refuse, to attend to its interest and to form practical maxims with a view to happiness" (5:61). But consistency with one's overall satisfaction is only a necessary condition for value, not a sufficient condition. A rational being does not "use reason merely as a tool for the satisfaction of his needs as a sensible being." He also has a "higher purpose," of accepting a principle that is a "practical law *a priori*"--a Categorical Imperative--and making it "the determining ground of the will without regard to possible objects of the faculty of desire" (5:62). Only choices that conform to moral law confer value, so it is "the moral law that first determines and makes possible the concept of the good" (5:64). Value is not determined by one's desires or feelings, nor is it a property that is passively perceived. Instead value is conferred by the choices of a being who acts upon rational principles of both prudence and morality. I do not claim that the ideas just summarized provide decisive arguments against the possibility of taking value as conceptually fundamental, but just that they summarize Kant's position.⁴

In some passages in *Groundwork*, including the first of Kant's two arguments that humanity must be the end in itself, it may sound as if Kant relies on fundamental claims about value. And it is true that Kant does not explicitly and forcefully articulate his atypical concept of value until the later *Critique of Practical Reason*. But despite his occasional terminological

lapses, Kant says even in *Groundwork* that value must depend on “law-making,” the activity of rational willing, “which determines all worth” (4:436). And all the major arguments of *Groundwork*, including the arguments for humanity as an end in itself, can be cast accurately in terms that do not rely on a conceptually fundamental claim about value.

It is quite possible to take Kant’s account of ends in a way that is consistent with the conceptual priority of rational choice. The most familiar type of ends are contingent or “subjective” ends, ends based on desires or feelings. It is not the case, on Kant’s story, that a desire is always a reason for action in itself. It is only a reason if one chooses to make it so. Ends, in other words, are always chosen, rather than one simply finding oneself with them. Once someone chooses to adopt an end, say the end of traveling to Egypt, Kant would then say that the end has value for that person. Kant describes the value of such contingent ends as a “relative” or “conditional” value. The value is relative, because traveling to Egypt only has value for an agent with a desire to travel to Egypt. The value is conditional in several ways. First, it is conditional in the same way it is relative—it only has value for someone on the condition of that person having the requisite desire. But it also is conditional on fitting with the overall package of an agent’s ends (so the end of traveling to Egypt has no value if traveling to Egypt will thwart the satisfaction of other, more important contingent ends), and its value also is conditional on being morally permissible (so if one must act immorally in order to be able to get to Egypt, the end has no value). In these ways, claims about the value of a contingent end are conceptually dependent on the fact that the end is rationally chosen.

A necessary end differs from contingent ends in that a necessary end provides a reason for everyone to act in certain ways, regardless of her particular desires. Contingent ends are, at least typically, states of affairs to be brought about, such as the state of affairs in which I visit

Egypt. But the only necessary end is rational nature, and it is not primarily a state of affairs to be brought about. It is a different kind of reason for action. Since rational nature is not something to be brought about to the maximum extent possible, we are not obligated to maximize the number of rational beings in the world, by having as many children as possible, nor are we obligated to maximize the number of rational choices that we make or maximize the extent to which we carefully ponder every choice. These consequences probably would follow from taking the value of rational nature as conceptually fundamental and then asking how to react to that value. But this is not Kant's approach. Instead, Kant's argument that rational nature is a necessary end, or end in itself, is meant to establish directly that we always, unconditionally, have a reason to treat rational nature in certain ways. That is, Kant does not at any point try to establish the conceptually prior value of rational nature and then derive duties regarding how to react to that value. Instead, he first argues that rational nature must be treated in certain ways, and any talk of the value of rational nature is just a shorthand for these requirements. Since a necessary end, or end in itself, is a reason that must carry weight for every agent in her deliberations, we can say that the value of a necessary end is absolute instead of relative, and unconditional instead of conditional on an agent's particular desires. Since no material reward can justify treating rational nature inappropriately, we can say that rational nature has an incomparably high value. But so far, this is not to specify the kinds of actions which are required by rational nature's status as an end in itself. To find these requirements, it is necessary to turn to Kant's argument for the legitimacy of the humanity formulation as a basic moral principle.

The Argument for the Humanity Formulation

A satisfactory reconstruction of Kant's argument for the humanity formulation has been elusive.

There are two general stages in Kant's overall argument. Kant first argues that if there is such a thing as a basic moral principle, then there must also be something that is an end in itself, because only an end in itself could ground a Categorical Imperative (4:427-28). This first stage is relatively clear. Then he tries to show that humanity or "rational nature" is the only satisfactory candidate for the position of end in itself (4:428-29). This second stage of the argument is remarkably cryptic, even in its most fundamental moves, so any efforts toward understanding the argument will necessarily involve significant reconstruction and filling in, rather than just interpretation. But the first, clearer, stage provides important clues for understanding the second, more difficult, argument that the end in itself must be rational nature.

Kant's arguments in the first two chapters of *Groundwork*, including the arguments regarding the end in itself, are not meant to stand independently of all everyday moral beliefs. Kant says in the preface to *Groundwork* that in the first two chapters his strategy is to "proceed analytically from common cognition [of morality] to the determination of its supreme principle," that is, to see what the content of the Categorical Imperative must be if it is to fit with basic everyday beliefs about the nature of morality (4:392). He says he will leave aside the project of establishing that there really are such things as moral principles -- that morality is not a "mere phantom of the brain" -- until chapter 3 (4:445). So, in arguing for the humanity formulation, Kant provisionally assumes that there are such things as basic moral principles, and points out that the only thing that could count as a moral principle or "supreme practical principle" is a Categorical Imperative, or a "universal practical law" that unconditionally demands compliance from everyone (4:428). Kant goes on to argue that if there is such a Categorical Imperative, then

there must also be something that is an end in itself.

Contingent or relative ends can not be the basis for a moral principle that necessarily applies to everyone, simply because such ends vary from person to person. Kant says

...their mere relation to a specially constituted faculty of desire on the part of the subject gives them their worth, which can therefore furnish no universal principles, no principles valid and necessary for all rational beings and also for every volition, that is, no practical laws. (4:428)

Therefore, if there are any genuine moral principles there also must be something that is an end in itself, an end that every agent must recognize as a reason for action regardless of her particular desires.⁵ But to say that there must be some end that applies necessarily to everyone if there is to be a Categorical Imperative is not yet to specify what that end must be.

Kant first simply asserts that a rational being is always an end in herself, in *Groundwork* 428, but then he offers two arguments for the claim. The first argument is also in *Groundwork* 428, and appears to be an argument by elimination. Kant examines and rejects three candidates for the position of end in itself. Since the argument is not compelling, I will focus on Kant's second argument instead.

This second argument for the claim that rational nature must be the end in itself, in *Groundwork* 428-29, is intriguing, but so compressed as to be largely mysterious. Kant certainly gives some signs that he takes it to be a sound deductive argument, using the word "therefore" (in German, "*also*") to mark the conclusion, but it is difficult to see exactly what the argument is. He is talking about what the content of the Categorical Imperative or "practical law" must be like, and says

The ground of this principle is: Rational nature exists as an end in itself. This is the way in which a human being necessarily conceives his own existence, and it is therefore so far a subjective principle of human actions. But it is also the way in which every other rational being conceives his existence, on the same rational ground which holds also for me; hence it

is at the same time an objective principle, from which, since it is a supreme practical ground, it must be possible to derive all laws for the will. The practical imperative will therefore be the following: Act in such a way that you treat humanity, whether in your own person or in any other person, always at the same time as an end, never merely as a means.

Several things are puzzling here. A satisfactory reconstruction would need to explain what reason there is to suppose that every agent “necessarily” conceives of herself as an end in herself, and what is involved in conceiving of herself in this way. And once it explained why each agent must conceive of herself that way (why treating humanity as an end in itself is a “subjective” principle), it would still need to explain why she should also conceive of other agents as ends in themselves too (why it is an “objective” principle). And a satisfactory reconstruction would need to do all this without relying on claims about the value of humanity. Claims about the incomparably high value of humanity should, strictly, be used only to capture conceptually prior ideas about the humanity formulation’s demands that humanity be treated as an end in itself.

The opening claim of the argument is already bewildering. Kant says that each rational agent necessarily takes her rational nature to be an end in itself, so treating her own rational nature as an end in itself is “so far a subjective principle of human action.” But Kant’s use of “subjective” here must be inconsistent with the definition he has just offered in 428, where he says “subjective” means something like “based on inclinations.” It would not be a *necessary* principle of action if it were based on an agent’s inclinations, since there could at least in theory be a rational agent who lacked the inclination that led her to act on the principle. Instead of using “subjective” here to mean “inclination-based,” I think he is using it in a more common, non-technical sense, to say that the principle has to do with one particular individual. The principle has as its content only the agent’s own rational nature, and applies to the agent’s own actions. That is, she treats her own rational nature as an end in itself, but so far there is no mention of

how she will regard others or how they will regard her. Nevertheless, the principle is still empirically false. Kant rightly acknowledges throughout his writings that it is empirically possible for an agent to fail to treat her own rational nature as an end in itself. Real human agents sometimes act immorally. So instead of being an empirical claim about how actual agents necessarily act, Kant's statement must be describing the manner in which agents are rationally required to act.

But even if I am correct so far, this is just to decipher the meaning of "subjective" in Kant's claim that treating one's own humanity as an end in itself is a necessary subjective principle. It is not yet to show what is involved in treating one's own humanity as an end in itself, nor why each rational agent should take herself as having reason to accept this requirement unconditionally.

Christine Korsgaard (1996, p. 119-124) provides a strategy that is helpful in deciphering Kant's argument for this "subjective principle" of action, and although some details of my reconstruction diverge from Korsgaard's, I will follow her basic strategy.⁶ She says "the argument is intended as a regress on the conditions" of the value of our contingent ends (Korsgaard, 1996, p. 120). So we should begin by asking what makes a contingent end worth pursuing. Following the Kantian reasoning outlined in the previous section, in the discussion of value and ends, we can see that just having a desire for something is not a sufficient condition for its value. Given the fundamental Kantian point that value depends on the rational choice of agents, an end that is contrary to rational requirements of prudence or of morality has no value. In more common language, not just everything one wants, nor even everything one chooses, has value. Since value is not prior to the rational choice of agents, it is incoherent to appeal to the supposed value of a contingent end in order to justify undermining one's own rational nature,

because “if you overturn the *source* of goodness of your end, neither your end nor the action that aims at it can possibly be good, and your action will not be fully rational” (Korsgaard, 1996, p. 123). The necessary condition of the value of any contingent end is that it must be an end set by some rational nature. And the rational nature that serves as the necessary condition of the value of contingent ends is “rational nature” in a strong, Kantian sense. It is the will of a being who accepts the force of rational principles of both prudence and morality. This rational nature is the end in itself.

A properly ordered will is the end in itself which should never be sacrificed for the sake of achieving contingent ends. If this properly ordered will, or fully rational nature, is the end in itself, then there are two ways to violate the subjective principle of not destroying one’s own rational nature. The most common way to give up one’s rational nature is to choose to act contrary to moral requirements, and so lose the commitment to morality that marks off rational wills. Consistently with this reading of the “subjective” component of the humanity formulation, Kant makes clear later that the humanity formulation includes a duty of moral self-perfection, to strive always to make self-legislated moral principles a sufficient reason for action (6:387, 6:392-93, 6:446-47). The second way to sacrifice one’s own rational nature is to destroy oneself or one’s minimal rational nature altogether. This kind of sacrifice -- cases of suicide or of placing oneself in situations that involve great risk of losing one’s life for the sake of satisfying contingent desires, or taking permanently mind-altering drugs or the like -- is probably what most naturally comes to mind when one thinks of sacrificing one’s rational nature, if one does not keep in mind that Kant’s idea of rational nature encompasses much more than minimal rationality. And Kant of course thinks that every rational agent has duties to herself to avoid these kinds of actions, duties based on the humanity formulation (6:422-23, 6:427-28). So, the

subjective principle that is suggested by the regress argument is to avoid sacrificing one's commitment to morality for the sake of satisfying contingent desires, and also to avoid sacrificing oneself or one's minimal rational powers altogether.

But this leaves a further step in reconstructing Kant's argument for the humanity formulation, namely establishing that treating humanity as an end in itself is an "objective principle" as well as a subjective one. The support Kant offers for the objective principle is that in the same way that every agent must think of her own rational nature as an end in itself...

it is also the way in which every other rational being conceives his own existence on the same rational ground which is valid also for me; hence it is at the same time an objective principle"(4:429).

In one sense, Kant's reasoning here is perfectly straightforward. The argument offered above in support of the "subjective principle" is perfectly general. It did not presuppose any particular desires, but rather showed that it is always illegitimate to appeal to the supposed value of contingent ends to justify compromising one's own rational will. So every rational agent has reason to treat her own fully rational nature as an end in itself. But Kant means the objective principle to establish more than this, namely that each of us must treat humanity as an end in itself, whether it is one's own humanity or someone else's. There is a large gap to be filled in the move from saying each agent must treat her own rational nature as an end in itself to saying that each agent must treat every rational nature as an end in itself.

Kant himself provides little clue about how to proceed, but an obvious strategy is suggested by the overall structure of *Groundwork*. Kant has supposed that if morality is to be more than a fiction, there must be a principle of morality that is binding on all rational beings. Two possible candidates for this universal principle are compatible with, and suggested by, the

fact that each rational agent must treat her own rational nature as an end in itself. One possible universal principle is: Each agent must treat her own fully rational nature as an end in itself, but may treat others' rational natures as expendable means to the satisfaction of her own desires. The other possible principle is: Each agent should treat her own fully rational nature and all other fully rational natures as ends in themselves, so should not trade any rational nature for the satisfaction of her own desires. If we limit Kant to morally neutral premises, he could provide no reason for thinking that the second principle is the correct one. But he does not limit himself to morally neutral premises. In these chapters of *Groundwork*, he is analyzing what morality must be like if there is any such thing as morality. While the first principle described above does provide an imperative that is in a sense universal, it is not universal in the sense required to count as a moral principle. It would give every agent a command that verbally had the same form-- "treat your own rational nature as an end in itself, so never sacrifice it for the sake of inclination"--but it would not be demanding that each agent treat exactly the same object(s) as deserving this special consideration. The common moral idea that is being analyzed demands an end that all moral beings can share, not one that will irreconcilably set them into conflict (5:28). If morality is not an illusion, it requires an end that can be shared by all agents, and that is what justifies the move from the "subjective principle" in the argument for the humanity formulation to the "objective principle" that one must treat fully rational nature as an end in itself wherever one finds it.

But what is involved in treating humanity as an end in itself? The subjective component of the humanity formulation forbids one to sacrifice one's own rational nature by choosing to place higher priority on contingent desires than on morality, or by altogether destroying oneself and so one's will, or by permanently impairing the basic functioning of the will. The

corresponding objective principle, because it must serve as a moral principle rather than a principle of strife, imposes roughly parallel requirements on the treatment of others' fully rational natures, so far as possible given the basic differences in the effects we can have on ourselves and on others. The objective principle demands that one not destroy others for the sake of satisfying one's own desires, and that one not permanently impair others' deliberative powers. Since you can not control the choices another person makes, or the principles she chooses to adopt, you can not strictly have an obligation to preserve others' commitments to morality. But you do have a "negative" duty not to tempt them to immorality (MM 394). And I think these requirements are all that Kant's argument in *Groundwork* 428-9 can establish.

And this is enough to reach the conclusions Kant wishes to reach in arguing for the basic humanity formulation, that humanity is an end in itself, and so should never be treated merely as a means. Humanity is an end in itself, or an objective end, because each agent is rationally required not to sacrifice her own or others' humanity, or fully rational will, for the sake of her own desires. This requirement applies to each rational agent, regardless of the desires she has. So humanity is an end that must be taken account of in action regardless of one's desires. That is what it is to be an objective end, or end in itself. The additional claim that rational nature should not be treated as a mere means emphasizes the way that one can fail to treat it as an end in itself. The way to violate the demand of treating humanity as an objective end is to undermine or destroy it for the sake of achieving desire-based ends. The language of "ends" and "means" is a little strained here, but that is not a feature unique to my reconstruction of Kant's argument. As teachers of introductory ethics classes know, it is fairly intuitive to describe some violations of the humanity formulation as cases of "using someone as a means" (deceiving someone, for example, or "using" someone in a romantic context to make another person jealous), but it is

more of a stretch to make the “treating as a means” label intuitively fit cases of suicide or non-beneficence. Kant, I think, is using the distinction between ends and means partly because it is a distinction that is readily available from the history of philosophy and seems at least roughly to capture the idea of giving the right (or wrong) kind of weight to something in one’s deliberations.

But even if everything I have said about the argument for the humanity formulation is correct, it does not settle all the duties regarding humanity. Surely the humanity formulation requires more than just not destroying oneself or others, and striving to regulate one’s choices with moral principles. But not all these other duties directly play a role in the argument for the humanity formulation. Instead, they are derived from the humanity formulation once it is established.

How Particular Duties Follow

It is no surprise that the argument for the humanity formulation does not itself specify all the duties entailed by the humanity formulation. Kant says “The present *Groundwork* aims only to seek out and establish the supreme principle of morality,” and he reserves for a later day the “application of that supreme principle to the whole system” of morals (4:392). He calls the whole system a “metaphysic of morals” (4:391). In the later work actually titled *The Metaphysics of Morals*, he says that

...a metaphysics of morals cannot dispense with principles of application, and we shall often have to take as our object the particular nature of man, which is known only by experience, in order to show what can be inferred from universal moral principles.
(6:217)

The Categorical Imperative is meant to apply necessarily to all possible rational beings. But to decide on most of the specific duties that follow from the general principle, one must take account of the circumstances and psychology of the kind of rational beings whom one is considering. To arrive at duties that apply to humans, one must consider human nature. If that strikes some readers as too contingent to be Kantian, I can only point out that it is Kant's own stated view.

Some commentators have suggested that respect or esteem for rational nature must play a key role in deriving specific duties from the humanity formulation (Wood, 199, p. 147-149, Hill, "Editor's Introduction" to Kant, 2002, p. 80-81). This is a sound approach, but it would be desirable to fill in more details related to this respect or esteem. It is not completely clear whether it is meant to be a feeling, an attitude, a disposition, or just a way of describing respectful actions. And if this respect is supposed to have rational nature as its object, then a more complete account would need to explain how rational nature gives rise to this feeling (or attitude, or disposition), and how particular duties are related to it. Luckily, Kant himself provides a description of a moral feeling that is well suited to play this role. It is the feeling of "*Achtung*" – of respect or reverence. As described earlier in this chapter (in defining the "capacity for morality"), the most common use Kant makes of the word "*Achtung*" is to refer to the feeling produced in finite rational beings by their recognition of the awesome force of moral law. But Kant says that people also can inspire this feeling of *Achtung*. In fact, in *Critique of Practical Reason* 76, which contains his most prolonged discussion of *Achtung* for moral law, he says that "[*Achtung*] is always directed to persons, never to things." *Achtung* for moral law and for persons is the same moral feeling, because the feature of a person that inspires a feeling of *Achtung* is her commitment to moral law. Only "uprightness of character" elicits *Achtung*,

because the person who displays good moral character provides an example of the power to rise above material circumstances. So Kant describes the feeling of *Achtung* as arising because a finite rational being “sees the holy elevated above itself and its frail nature,” and Kant is consistent throughout his writings on ethics that people who are committed to moral law inspire the feeling of *Achtung* by their example (6:464, 4:435, 4:440). This moral feeling of *Achtung* for persons is well suited to play a theoretical role in the transition from the humanity formulation as a basic moral principle to the particular duties that follow from the principle.

The most obvious category of duty that is related to the feeling of *Achtung* is the category of duties of respect (*Achtung*) for other people. Admittedly, there can be no duty to feel respect, because one either has a feeling or not, and we can only have duties to do what is within our power (6:449). But we do have duties to act in ways that embody respect, or in other words to act in the manner of someone who actually feels respect. So Kant says every person has a duty “to acknowledge, in a practical way, the dignity of humanity in every other man. Hence there rests on him a duty regarding the respect [*Achtung*] that must be shown to every other man” (6:462). The feeling of *Achtung* for other rational beings would lead to a recognition that others can be as important as oneself, and so would tend to quash a feeling of arrogance. And it would lead one not to condemn or ridicule others, so as not to drag them into a position lower than they deserve. This fits with Kant’s description of the vices opposed to *Achtung* for others, the vices of arrogance, defamation, and ridicule (6:465-68).

Kant also describes several duties that would most naturally be described as duties of respect for oneself. Kant hesitates to say that we have a duty to respect ourselves, but again he means that we can not properly say we have a duty to have a feeling of *Achtung*. He says, “...it is not correct to say that a man has a duty of self-esteem; it must rather be said that that the law

within him inevitably forces from him respect for his own being, and this feeling (which is of a special kind) is the basis of certain duties...” (6:402-403). So Kant means to rule out a duty to have a feeling of respect for oneself, but he does not rule out that one can have duties to perform the kinds of actions that express respect for oneself. A person’s duty to avoid servility is a duty not to act contrary to “his consciousness of his dignity as a rational man, and he should not disavow the moral self-esteem of such a being” (6:435). Each of us also has a duty to avoid avarice, and this avarice consists of “...restricting one’s own enjoyment of the means to good living so narrowly as to leave one’s own true needs unsatisfied” (6:432). One of the ways in which we express *Achtung* and esteem for the will of a being committed to morality is to take satisfaction in seeing such a being made happy, not because material reward is the motive for moral commitment, but because we inevitably see virtue as worthiness, and as material beings we see worthiness as worthiness to be happy.

Achtung plays a similar role in the derivation of the duty to aid others in the pursuit of their ends. The argument for the basic humanity formulation by itself does not establish that one must aid others in achieving their ends in general. It does seem to show that one ought to aid others when their survival or their powers of rationality are threatened, since one ought never to sacrifice any being’s rational will for the sake of satisfying one’s own desires. And sometimes Kant does speak of the duty to aid others who face great danger or hardship (4:423, 6:453). But Kant also has in mind a more general duty of “making others’ happiness one’s end...” (6:452) or “to further the ends of others” (4:430). The best way to account for this more general duty of aiding others in the pursuit of their ends is through the feeling of *Achtung*. The feeling of the incomparable worth of other rational beings committed to morality combats my natural tendency to arrogance and makes me aware that my own contingent ends are not uniquely important. Since

another person's proper willing makes her ends worth pursuing as well, I ought to acknowledge that they are not worthless. A way to do this is to provide some aid for her in pursuing her ends, if doing so is not too great an infringement on my own pursuit of my ends.

The duties to oneself of natural self-perfection also seem best derived from the humanity formulation by employing the feeling of *Achtung*. This captures the spirit of Kant's claim that to develop one's natural abilities is to make oneself worthy of one's own humanity (6:392, 6:387). The end-setting of a properly ordered will results in ends that are worth pursuing. Some of these ends are based on inclination, others are moral ends given unconditionally by reason. To develop one's abilities allows one to seek a wider range of contingent ends, and also allows one to seek moral ends in a wider variety of ways. One can have a good will without possessing the ability to achieve a wide variety of ends, but the feeling of *Achtung* produced by a (fully) rational will inspires us to make it possible for ourselves to set and achieve a wide range of ends. So Kant says, "... there is also bound up with the end of humanity in our own person the rational will, and so the duty, to make ourselves worthy of humanity by culture in general, by procuring or promoting the capacity to realize all sorts of possible ends..." (6:392). This also fits with the spirit of Kant's statement in *Groundwork* 430 that although failing to develop one's talents does not "conflict" with "humanity in our own person" and is consistent with the "maintenance of this end" it also does not "harmonize with this end." Failing to develop one's abilities does not literally destroy one's rational will, but it is not consistent with fully accepting and acting on the feeling of respect and worthiness to be happy that a rational will inspires.

Although these are not all the duties that Kant discusses, they are sufficient to suggest the general pattern of the derivation of particular duties from the humanity formulation, and to demonstrate the usefulness of the concept of *Achtung*.

Final Thoughts

Earlier, I set aside the issue of exactly which sort of “rational nature” is best taken as the end in itself, on the grounds that the answer would be clearer after looking at Kant’s arguments. The arguments have suggested that what Kant means to identify as the end in itself is rational nature in a quite strong sense, as the entire rational nature of a being, but only on the condition that the being is committed to accepting the force of moral demands. So just the power to set ends is not an end in itself, nor is just the power to legislate moral principles, nor the (unrealized) capacity to act on moral principles.

The “regress argument” for the subjective component of the humanity formulation leads to this conclusion, since a contingent end only has value on condition of being set by a rational being who accepts principles of prudence and morality. This also fits with the claim of the opening paragraphs of *Groundwork*, that a good will is the only thing unconditionally good, and the necessary condition of all other value (4:393-94).

The strategy I offered for deriving particular duties from the humanity formulation provides additional support for the “good will” reading of the humanity formulation. Other commentators have suggested that a general feeling of respect or esteem for rational nature is the key in moving from the general moral principle to more specific duties, and this is plausible enough. But an account of this feeling is not strongly connected to Kant’s texts, unless one takes the feeling to be the same *Achtung* that one feels for moral law. Taking the end in itself to be a good will draws a deep connection between the content of the Categorical Imperative and the effect of the Categorical Imperative on the moral agent who is subject to it. The Categorical

Imperative gives rise to a feeling of *Achtung* in moral agents who are aware of its force. If the rational nature that is an end in itself is a good will, then rational nature on its own accord also gives rise to the feeling of *Achtung*, because a good will is an example of the Categorical Imperative's power to outweigh contingent desires. The humanity formulation does not just command each agent to treat something as an end in itself, but more profoundly says to treat as an end in itself the kind of will that arouses the same deep moral feeling of *Achtung* as the moral law itself does.

Kant's presentation of humanity as an ideal which we ought to pursue also supports taking humanity as a good will. In many texts, Kant says that we ought to strive toward moral perfection, though it is a goal we will never fully achieve (3:384, 4:569, 6:61-63, 6:183, 6:387, 6:392-393, 6:446). And he frequently refers to this goal or ideal as "humanity." Although we will always be imperfect, because we are subject to temptations, the striving itself is the form of moral goodness that is possible for humans. The very concept that humanity is an ideal to be pursued seems to rule out identifying the humanity that is an end in itself as something that every minimally rational agent possesses. If everyone already possesses it, it could not be something to work toward. If it seems odd to identify "humanity" as something to work toward, it may help to keep in mind some points about ordinary language. In English, the injunction to "be a man" is familiar, and so is the less sexist demand to be a little more human. In German, at least the German of roughly Kant's time, one finds the statement in Mozart's *Die Zauberflöte*, that the character Zarastro must undergo trials to learn to be a man ("ein Mensch zu sein"). The familiar Yiddish use of "Mensch" to mean not just any human being, but a decent, reliable, or upstanding human fits even more nicely.

The main obstacle to accepting good will as the end in itself is the understandable

concern that this will require judgments of others' moral character, and will license mistreatment of people whom one deems immoral. But the distinction already described above, between basic moral principles and the application of those principles, provides a non-ad hoc defense against the suspicion of excessive moralism. At the level of moral theory, it is true that the end in itself is good will, or rational nature in a quite strong sense. But at the level of application to human conditions, we always (or almost always) have reason to treat other humans as ends in themselves, even if not all humans fully deserve this treatment. Kant says this himself. In the passages in *The Metaphysics of Morals* where he describes duties of respect for others (the same passages quoted above, in the previous section), Kant says that we must treat *all* other humans with respect, but he quickly adds that this is *not* because they all necessarily are worthy of respect. He is quite explicit that we should respect the vicious man "even though by his deeds he makes himself unworthy of it" (6:463).

Why should we treat him with respect if he is unworthy of respect? Kant gives three reasons. The first is that to treat a human with contempt weakens her belief that she can improve herself and so discourages her from attempting to better adhere to the dictates of practical reason (6:463-464, 466). A second reason for treating all humans with respect, even though only some deserve it, also depends on basic human psychological tendencies. Kant says that to treat any human with disrespect lessens our respect for all humans, "...so as finally to cast a shadow of worthlessness over our race itself, making misanthropy (shying away from men) or contempt the prevalent cast of mind..." (6:466). Finally, the most fundamental reason for treating everyone as an end in herself, even though not everyone actually deserves such treatment, is that we are not reliable judges of others' moral character. We can not even be sure what reasons an agent acts on in a particular case, let alone whether she embraces a higher-level principle of only acting in

ways that are morally permitted. Kant maintains that it is impossible to know with certainty whether a right action has been performed because of its rightness, or from some inclination (4:407). He is even more explicit in *Religion* that, although we can observe an agent perform impermissible actions, “...we cannot observe maxims, we cannot do so unproblematically even within ourselves; hence the judgment that an agent is an evil human being can not reliably be based on experience” (6:20, also 6:47-48, 67, 71). Avoiding judgments about others’ overall moral character is all the more important, given the innate human tendency to elevate one’s own worth in comparison to others’. Kant consistently attributes to humankind a self-love and self-conceit that leads to competition, both in the form of “an unjust desire to acquire superiority for oneself over others” and “an inclination to gain worth in the opinion of others.”⁷ So it is not just technically impossible to achieve absolute certainty in our judgments about character, it also is very likely that such judgments will be distorted. Kant, despite his reputation for being out of touch with human nature, provides three plausible psychological claims that provide reasons for us to treat other humans as ends in themselves, even if not everyone has earned this treatment.

The point is not trivial in Kant’s moral system. He thinks that we humans must treat the obviously virtuous and the apparently immoral in mostly the same ways, but he can not mean that this is because they really deserve the same treatment. Kant maintains that we are not in a position to make judgments of others, character, but he does think that God could make such judgments (6:48, 6:76-77, 5:123-4). In fact, that is the basis of Kant’s arguments that one rationally can believe in a supreme being – only such a being can judge others’ worthiness to be happy, and apportion their rewards to their worthiness.

If an unpalatable flavor of moralism still clings to the reading of “rational nature” I have proposed, perhaps some large-scale considerations may help. While it is both fashionable and

laudable to speak of basic human rights and inalienable dignity, the point of such talk is usually to place requirements on the ways that governments or other institutions must treat people. The reading I have proposed does not undermine such demands, and in fact provides a strong philosophical basis for them. The idea that the morally worst sort of people may not really have the same moral status as the morally best is compatible with the idea that we are not in a good position to label anyone as fundamentally inferior or less deserving of respect. This duality of thought is hardly a novelty. It is familiar to billions of believers in a supreme God, and also to believers in the principle of treating a defendant in a legal case as innocent until proven guilty. If there is anything odd about the picture, the oddity should be balanced against the competing picture, that no amount of immorality can ever tarnish the inextinguishable glow that accompanies the power of choice or the capacity for morality, even if the capacity is never realized and the choices are routinely monstrous. The position that we all ought to strive toward an ideal of moral goodness is at least less peculiar than that.

RICHARD DEAN

¹ All direct quotations from Kant's works will be from the translations cited in the bibliography, although page references will be to the standard Royal Prussian Academy edition of Kant's works.

² Of course, I am not even pretending to summarize Kant's arguments for this claim. See chapter 3 of *Groundwork*. A good secondary source on this is Hill, 1992, p. 97-122.

³ See Hill, 1992, p. 40-41, Herman, 1993, p. 238, O'Neill, 1989, p. 138, Paton, 1947, p. 177, Ross, 1954, p. 52-53.

⁴ Although the reading I give here is more or less standard, for contrary views see Herman, 1993, p. 208-210, and Guyer, 2000 p. 2, p. 96-171.

⁵ Kant summarizes this argument in basically the same way in 6:381.

⁶ The main difference is that Korsgaard takes the argument to establish that the mere power to set ends is the end in itself.

⁷ Both quotations are from 6:27. See also 7:272 and 6:465, and various of Kant's essays of the Critical and Post-Critical period, most notably "Perpetual Peace," "Idea for a Universal History with a Cosmopolitan Purpose," and "The Contest of Faculties."