Oxford Handbooks Online

Abstract and Keywords

This article focuses on how certain basic normative questions are addressed by Kant and various contemporary Kantians who interpret and extend Kant's theory. Kant's ethical writings are open to different interpretations, and the literature devoted to interpreting and extending his ideas is vast, diverse, and of mixed quality. When illustrating the action-guiding use of Kant's basic principles, this article limits its discussion by concentrating primarily on cases of beneficence and mutual aid rather than addressing a full range of moral problems. Its primary aim is to call attention to a variety of different ways in which contemporary Kantians are attempting to develop Kant's normative ethics. A secondary aim is to illustrate these different developments by relating them to a particular task that any normative ethical theory needs to address.

Keywords: Kant's theory, ethical writings, mutual aid, moral problems, normative ethics

Kant is often studied for his contribution to debates that today might be classified as metaethical,[1] but my concern here is with questions of normative ethics. In particular, I want to focus on how certain basic normative questions are addressed by Kant and various contemporary Kantians who interpret and extend Kant's theory. The main questions are familiar ones: Are there many basic moral principles or only one? How are we to articulate and interpret the basic principle or principles? How do they, or does it, function as an action-guide for particular moral problems? Are the basic Kantian principles really (as Kant thought) the norms underlying common moral judgments?

My discussion will be wide-ranging, but still limited in several ways. Kant's ethical writings are open to different interpretations, and the literature devoted to interpreting and extending his ideas is vast, diverse, and of mixed quality. Although I draw from several prominent contemporary Kantians, I cannot survey all of the good literature in this area. When illustrating the action-guiding use of Kant's basic principles, I limit my discussion by concentrating primarily on cases of beneficence and mutual aid rather than addressing a full range of moral problems. Because I review many different interpretations, my description and illustration of each, including the ones that I find most promising, must be quite brief. My primary aim is to call attention to a variety of different ways in which contemporary Kantians are attempting to develop Kant's normative ethics. I hope (p. 481) that this will encourage further investigation and development of the views only sketched here.

A secondary aim is to illustrate these different developments by relating them to a particular task that any normative ethical theory needs to address. This is to determine, at least in broad terms, when and why we are morally required to help others. Kant's theory is justly famous for its insistence that pursuit of happiness, for both oneself and others, is constrained by moral requirements of justice and respect for human dignity. Kant also insists, however, that it is categorically imperative for us to make it our maxim to promote the happiness of others. It is currently a matter of controversy how much latitude this requirement is supposed to leave us to pursue nonobligatory projects of our own.[2] Focusing on beneficence to illustrate various contemporary developments in Kantian ethical theory should bring out similarities and differences among them. It may also reveal some of their relative strengths and weaknesses.

1. Willing Maxims as Universal Laws

1.1. The Formulas of Universal Law

Kant's interpreters have most often taken as his primary action-guiding principle the first formulation of the Categorical Imperative, the formula of universal law. The initial expression of this in *Groundwork* I is: "I ought never to act in such a way *that I could not also will that my maxim should become a universal law*" (Kant, 2002, p. 203 [4:402]).[3] Kant's first formula of the Categorical Imperative in *Groundwork* II is commonly thought to express the same idea: "Act only on that maxim by which you can at the same time will that it should become a universal law" (p. 222 [4:421]). Kant then offers a variant of the formula: "Act as though the maxim of your action were to become by your will a universal law of nature" (p. 222 [4:421]). Following tradition, we can label the first two quotations as expressions of the *formula of universal law (FUL)* and the third as an expression of the *formula of universal law of nature (FULN)*. Kant repeats these formulas with variations in wording throughout his ethical writings.[4]

All of these are supposed to express the only principle that can be, in the strictest sense, a "categorical imperative," and so it is called a formulation of *the Categorical Imperative* (Kant, 2002, p. 214 [4:413]). This implies that it expresses in the form of an imperative an unconditional basic requirement of practical (p. 482) reason. An *imperative*, in Kant's technical terminology, expresses an objective principle as a *constraint* on imperfectly rational persons. An *objective principle* is one that any *fully rational* person would follow— but human beings (who are imperfectly rational) might not (p. 214 [4:401]). Kant later states several other formulas of the Categorical Imperative, but he claims that these are simply different ways of expressing fundamentally the same principle—the supreme principle of morality (p. 237 [4:436]). He sometimes refers to more specific moral principles as "categorical imperatives." This may cause confusion, because he says explicitly that there can only be one categorical imperative (p. 222 [4:421]). We can understand his view as consistent if we take the Categorical Imperative, expressed in various formulas, as the only imperative that is categorical *in a strict sense* but then add that more specific moral principles can be called "categorical imperatives" *in an extended sense* if they are derived from the Categorical Imperative and hold without exception.

Commentators differ significantly about how Kant's formulas are supposed to work as action-guides, but some points seem clear enough.

First, to determine whether a proposed act would be right, we must identify "the maxim" of the act. Maxims are "subjective principles" on which we act (Kant, 2002, p. 202 n. [4:401 n.]).[5] The fullest statements of them describe the act, its purpose, and the underlying reason (at least as the agent understands these).[6] They can be expressed as policy statements with the form "In conditions C, I shall do X in order to E from the motive M." They may be very general, such as "I shall always do what best serves my own interests," or quite specific, such as "When in need and aware that I cannot repay loans, I shall borrow money anyway with a (false) promise to repay."[7] Some maxims are morally bad, some are good, and many are morally indifferent. Kant offers the universal law

formulas as a way of testing whether acting on our proposed maxim would be wrong.[8] Of course, it may actually be rare that we have in mind an explicitly articulated maxim when we act. If, however, we sincerely question whether a proposed act would be morally permissible, we can reflect on what we are about to do, and why, and from this try to construct the relevant maxim to test by the universal law formulas. How exactly this should be done remains a problem, but certain guidelines seem implicit in the aim to deliberate conscientiously. For example, our maxim can only refer to facts of which we are aware, and it should reflect honestly our beliefs about what is morally relevant in the situation. Apart from this, it should not include details that are irrelevant to our policy, purpose, and reasons for acting as we propose to do.[9] If our first attempts to articulate our maxim fail the test, it may be that, on honest refection, we can rephrase the maxim in way that more aptly describes what we propose to do, our purpose, our reasons, and the limits of the policy that we mean to endorse.[10] Since the practical purpose of the formula is to guide conscientious deliberation about whether it is permissible to act as we (p. 483) are inclined to do, it would be both dishonest and self-defeating to try to rig the statement of our maxim so that Kant's tests will "justify" what we really believe is wrong. Philosophers sympathetic to Kant's project also may be suspected of rigging maxims in a dishonest and self-serving way if, *whenever* faced with counterexamples, they keep redescribing the maxim until Kant's texts yield the intuitive result that they want. If that is necessary, the formula is not really serving as a decision guide.

Second, Kant distinguishes between two kinds of maxims that fail the test posed in the universal law formulas. Some maxims cannot be *conceived* as universal law[11] without contradiction; others can be, but cannot be consistently *willed* as universal law. Thus, in effect, our maxims must meet two requirements. The first is that we must be able to *conceive* our maxim as a universal law without contradiction. If we cannot, then it is wrong to act on the maxim. If we can conceive our maxim as a universal law without contradiction, we must then still ask whether we can *will* our maxim as a universal law. Maxims that satisfy both requirements are supposed to be permissible to act on; those that fail either requirement are supposed to be wrong to act on. In the *Groundwork*, Kant gives two examples of maxims that fail the first test and two examples of maxims that fail the second test.[12]

Third, although Kant presents these requirements as tests of the *permissibility* of acting on a proposed maxim, they could lead us to the conclusion that we have a positive *duty* to act in certain ways. If we wonder whether it is also duty to act on a given permissible maxim, then we must consider what our maxim would be if we chose to do otherwise. If we could not consistently conceive and will that alternative maxim as a universal law, then it is wrong to act on the alternative maxim. Whenever it is wrong to act on the only alternative to a permissible maxim, it is a positive duty to act on the permissible maxim. There may, of course, be many alternatives, but at least we can say that it is a positive duty to act on the initial permissible maxim, unless at least one of the alternatives can be conceived and willed as a universal law. Consider, for example, a maxim to aid persons in distress, at least when one can at little cost to oneself or others.[13] Presumably, this can be conceived and willed as a universal law, and so it is permissible to act on this maxim.

When confronted with a particular case, we need to consider what our maxim would be if we did not act on this maxim to help the distressed person. There are obviously many other things we could do instead, but presumably the maxims we would be acting on if we did not help would need to make reference to the morally salient fact that we would be refusing to help someone in distress even though we could easily aid the person without harm to ourselves or others.[14] If these alternative maxims cannot be conceived and willed as universal laws, then not to give aid in the circumstances would be wrong, and so giving aid would be a positive duty.

(p. 484) 1.2. Questions of Interpretation

This is not the place to try to evaluate all interpretations with regard to either their fidelity to texts or their plausibility as moral standards, but here are a few of the variations on Kant's theme.

A preliminary question concerns the relation between FUL and FULN. Do both of these offer procedures for testing the morality of maxims? If so, do they propose exactly the same test or different ones? On one reading, the more abstract FUL is not a practical action-guide by itself but expresses an idea that can be applied to human conditions only when "universal law" is replaced with the more specific concept "universal laws of nature."[15] On another reading, FULN simply specifies what was implicitly intended in FUL, and so the two formulas offer exactly the same test for maxims. A third possibility is that the FUL and FULN offer slightly different tests. In one version, FUL is concerned with *laws of freedom*, while FULN is concerned with *teleological laws of nature.* Thus the first asks us whether we can will that everyone freely choose to act on our maxims, and the second asks us whether we could consistently conceive and will our maxims as teleological laws in a harmonious system of natural purposes.[16] A more promising variation, in my view, is this: FUL asks us whether we can consistently conceive and will our maxim as *permissible* for everyone to act on,[17] and FULN asks us to whether we can consistently conceive and (if we had the power) will that, as if by a law of nature, everyone adopts and acts on the maxim.[18]

Other questions arise about what might, in a relevant sense, prevent us from conceiving a maxim as a law, or law of nature, for everyone. Most obviously, if everyone's acting (or being permitted to act) on a given maxim is logically impossible, then the maxim cannot be conceived as a universal law in the relevant sense. Kant's examples, however, suggest that they be ruled out if, *assuming some general background facts* about the human condition, it is logically impossible for everyone to act on them.[19] On another interpretation, they are excluded if it is logically impossible for a harmonious system of nature to include everyone's acting in the manner *and for the purpose* indicated in the maxim.[20] A more promising suggestion, perhaps, is that maxims cannot be conceived as universal laws (or laws of nature) if it is logically impossible to will simultaneously the following set of intentions: (1) to act on the maxim, (2) to bring it about (if one had the power) that everyone else act on the maxim, and (3) the normal foreseeable

consequences of everyone's acting on the maxim.[21] This is a particular version of the idea that we cannot conceive of a maxim without contradiction if the maxim "would be self-defeating if universalized: [Our] action would become ineffectual for [our] purpose if everyone (tried to) use it for that purpose" (Korsgaard, 1996b, p. 78).

There are also different ways of understanding the relevant impossibility of *willing* maxims that can be conceived as universal laws. I cannot review them all, but here is a sample. The least plausible, but all too common, idea is this means (p. 485) simply *being unwilling, for any reason, to choose* that our maxim be adopted by everyone. If, for example, racial bigots, because of their prejudice, would not be willing to for everyone to work for racial equality, then they could argue that by Kant's principle they would be wrong to do so. Even worse, they could argue that because they are willing for everyone to adhere to their policy of strict racial segregation, it is morally permissible for them to pursue it.[22] Equally, but less obviously, implausible is the assumption that *any contingent inability* (in our nature or circumstances) to will our maxims as universal law means that we would be wrong to act on them. Especially when we consider rather specific maxims, this would mean that all sorts of morally irrelevant factors would rule out actions that are quite innocent. If there were not enough of some trivial commercial product for everyone to buy it, the test would apparently show that it is wrong for anyone to buy the product. If, because of dizziness, some people cannot climb high ladders, it would seem that no one should.

Now, by reformulating the maxims and adding ad hoc stipulations to the test, we might circumvent these and other particular counterexamples; but arguably the interpretation in question is mistaken in principle. What should be relevant under the universal law tests is whether we can *rationally* endorse a maxim as a universal law.[23] The fact that not everyone *can* act in certain ways is often no reason for others to refrain, and all sorts of morally irrelevant quirks and prejudices may determine whether we are willing to endorse policies for everyone. What we need to consider is whether endorsing our maxims as universal laws is *contrary to reason* in some relevant way.[24] Obviously, maxims that cannot be *conceived* as universal laws without contradiction cannot be *rationally willed* as universal law, but beyond this, the relevant standards of rational willing are more controversial. A standard of logical consistency and coherence among one's intentions is unlikely to be sufficient by itself to generate appropriate results from the universal law tests. In a Kantian theory, however, these standards should not be intuitive moral norms that have no basis at all in Kant's moral theory, at least if we accept the common view that all other moral norms are derivative, in some sense, from Kant's basic moral principles.[25] In applying Kant's formulas, however, we should be able to rely on whatever specific moral norms we have already confirmed, or could confirm (on Kantian grounds) as rational requirements.[26]

Readers will need to assess the various interpretations for themselves, but a particularly important further controversy should be mentioned. In assessing whether we can *reasonably* will our maxims as universal laws, can we rely on the other formulas of the Categorical Imperative? To try to do so would be useless if, as some think, the later

formulas add nothing new and action-guiding beyond what the universal law formulas say.[27] If, however, as others claim, the later formulas articulate basic moral standards apart from (or at least not explicit in) the universal law formulas, then they would provide resources for arguments that certain maxims cannot be reasonably willed as universal laws. For example, if we (p. 486) exclude the later formulas, a rich miser who abhors charity for anyone might argue that he can will his maxim of refusing to help the needy as a universal law because he is prepared to die before accepting charity. If, however, we accept as a rational requirement that we treat humanity in each person as an end-in-itself, then arguably the miser could not reasonably endorse his "no charity" policy as a universal law.[28] Kantians take different positions on whether later formulas can supplement the universal law formulas, depending on their different interpretations of the textual evidence regarding Kant's intentions and their philosophical judgment as to what makes most sense.

1.3. Illustration: Helping Others in Distress

Kant's project in the *Groundwork* and *Critique of Practical Reason*, where the formula of universal law figures prominently, was not primarily to explain and illustrate how to apply his formulas to particular problems. In both works, and especially the latter, more attention is devoted to discovering and defending the basic presuppositions of the use of practical reason—finding its constitutive principles and their relation to freedom of the will (in several senses). *The Metaphysics of Morals* is the work where Kant turns explicitly to the task of working out intermediate principles for guiding ethical judgment in various areas of human life, and here the universal law formulas play a more modest role. In fact the idea of humanity as an end-in-itself is what Kant's arguments most often appeal to. Nevertheless, in the *Groundwork* Kant expresses great confidence that the universal law formulas can serve as guide to moral judgment, enabling us to distinguish right from wrong in every case (Kant, 2002, pp. 204–206 [4:403–4], pp. 224–226 [4:424–425]). Moreover, he suggests that, although ordinary people can become confused, they implicitly respect and rely on the universal law formulas as a standard (pp. 204–207 [4:403–405], 1997, p. 7 n. [5:8 n.]). Not surprisingly, not all readers have shared Kant's confidence in the formulas as an action-guide. For centuries now, critics have pointed out problems in applying them, and Kant's supporters have developed subtle defenses of his basic idea beyond anything that we can find explicitly in Kant's texts.

Let us consider how the universal law formulas might guide our moral reflection about helping others. First, we must acknowledge that many ways of helping others are morally impermissible. Justice, respect for others, and certain "perfect duties to oneself" set limits to what we may do to aid and promote the happiness of others. These constraints call for justification, by the either universal law formulas or later formulas, but let us assume for now that these restrictions can be justified. The general question is, if we fulfill our duties of justice, respect, (p. 487) and so on, what more must we do for others? To apply the universal law tests, we must look at the problem initially from the point of view of a particular agent in a particular context. Suppose, then, you are inclined to refuse an appeal to give to famine relief, but you wonder if this is morally permissible. The first thing to consider is what maxim honestly and accurately reflects what you propose to do. Cases will vary, but suppose you are inclined not to help anyone beyond what justice demands, though you could easily make a gift that would help to relieve someone's distress (Kant, 2002, pp. 224–225 [4:424–425]). Your maxim might be "From self-interest I will always refuse to help others in need except when they have a right to my aid, even though I could easily help without significant harm to myself or others." Presumably, you can consistently conceive a world in which everyone adopts and acts on this maxim. If everyone did so, however, and you should fall into dire need, by your policy others would help you only if you could demand help as a matter of justice (your rights). This, in most cases, would be contrary to your self-interest, and, by hypothesis, your purpose in adopting your maxim was to advance your self-interest. So willing that

everyone adopt and act on your maxim would be willing a situation incompatible with your aim in adopting your maxim. You cannot rationally will both, and so, it seems, you would be wrong to act on your maxim.

Suppose now that you are an unusually wealthy, secure, independence-loving miser. You might object to the previous argument, saying that you would prefer to die rather than accept aid from others. Then, you might argue, you can consistently act on your personal preferences (expressed in your maxim) while also willing your maxim for yourself *and everyone else*, for in this case your purpose in adopting your maxim was to satisfy a personal preference that you *and everyone else* refuse to give to charity. To meet this objection, one might argue that you cannot *rationally* chose that you would forego the means to your survival if you should happen to fall into dire need.[29] Arguably, the issue here is not the odds of your falling into dire need but rather what you can rationally will for the *possible* circumstance in which it happens, whatever the odds against it. As a *rational* person, one might argue, you necessarily value your existence as a rational agent over inclination-based preferences, and therefore you must treat the means to your survival as having a value of higher priority than your disinclination to accept charity. If so, it would be contrary to reason to will your maxim as one for everyone to follow. The upshot is that it is wrong for you to act on your maxim; and so unless you can honestly say that your refusal to give aid is based on some other, morally acceptable ground, you have a positive duty to give aid on the occasion in question. Furthermore, assuming you will have normal opportunities and abilities in the future, arguably you can ensure that you will not act on your impermissible maxim (or similar bad maxims) in the future only if you make it your principle generally to help others in need when you can at little cost to (p. 488) yourself or others.[30] None of this implies, however, that you must try to maximize happiness or to work for the happiness of others *whenever you have no conflicting duties.*[31]

1.4. Problems and Doubts

The most persistent worry about Kant's universal law formulas is that they often seem to lead to intuitively unacceptable conclusions. They apparently condemn some maxims that we regard as innocent and fail to condemn maxims that we regard as immoral. Frequently, revising the description of the maxim leads to more acceptable results, but there seems no principled way to tell before applying the tests what the "correct" description should be.[32] Even if we can always find some apt maxim description that allows us to reach common-sense conclusions, we are not really being guided by the formulas if we need to rely on our understanding of the right conclusion in order to find the best statement of the maxim. Even if the tests appropriately show that acting on certain maxims is *wrong* and acting on others is *permissible*, they can generate a *positive duty* to do something only indirectly by showing that it is wrong not to do it. Since there are usually many ways of *not doing* something, we would need to identify and test all of the many maxims that we might follow if we did not do as the alleged positive duty would prescribe. Defenders of Kant's universal law formulas have devised many subtle supplements to these as maxim-testing procedures, but arguably the proposals are merely ad hoc devices to patch up a flawed procedure or else they amount to an admission that the universal law formulas alone are not sufficient for determining particular moral requirements.

Another recurrent concern about treating the universal law formulas as the sole, or primary, moral action-guide is that the recommended test procedure, by itself, seems not to reflect what is most central to moral deliberation. As many Kantians now admit, even if the universal law formulas can flag certain maxims as morally wrong, or at least suspect, they do not adequately explain *why* acting on those maxims is wrong. What is wrong with slavery, for example, is not adequately explained by saying that it is impossible for everyone to act the maxim of a would-be slave-owner. It may be that those who rob banks and commit murder cannot consistently will both their maxim and that everyone act on that maxim, but this inconsistency seems at best only part of the story why such acts are wrong. Although Kant's formulas do not work well as precise decision procedures for particular moral cases, they should at least reflect, in a general way, essential features of a Kantian moral perspective for thinking about particular problems. Taken by themselves, the universal law formulas seem inadequate for this purpose. Arguably, Kant himself thought that the full import of the moral law becomes clear only when all of the formulas are fully taken into account.

(p. 489) 2. Treating Humanity as an End-in-Itself

2.1. The Formula of Humanity

Kant's next formulation of the Categorical Imperative is this: *"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"* (Kant, 2002, pp. 229–230 [4:429]).[33] He elaborates this formula in various ways. He says both that *persons* are ends-in-themselves and that *humanity, or rational nature*, in persons is an end-in-itself.[34] An *end*, in Kant's broad sense, is "what serves the will as the objective ground of its self-determining"; that is, it gives us a *reason* to do or refrain from doing various things (p. 228 [4:427]). Ordinarily we think of an end as something we aim to achieve or promote, but Kant says that an end-in-itself in his sense is a "self-sufficient" or independently existing end, rather than an end to be produced (p. 238 [4:437]). An end-in-itself is necessarily an end for every rational being: That is, its existence is an objective reason for doing or refraining from certain acts, independently of our inclinations. To value rational persons as ends, we must not use them for ends that, in some sense, they cannot share (pp. 230–231 [4:430]). Kant adds that persons, conceived as members of a kingdom of ends, have a *dignity*, which is grounded in their *autonomy* of will (pp. 234–237 [4:434–436]).[35] Dignity is an "unconditional and incomparable worth," above all *price* and "without equivalent." Thus dignity is a value that is independent of a person's social status and utility, and it is not to be exchanged for anything with merely conditional value.[36] Negatively, it is a value against which we must never act; and yet, positively, we take the conception of humanity as an end-in-itself fully to heart only if we try to make the (permissible) ends of others our own (pp. 230–231 [4:430], pp. 237–238 [4:437–438]).[37]

2.2. Questions of Interpretation

The humanity formula, like the previous ones, has been interpreted in a variety of ways. One question is whether the formula has any independent action-guiding content. Obviously, it would not if the test for our treating humanity as an end were simply that we could will the maxim of such treatment as universal law.[38] Similarly, if the formula meant no more than "Respect persons' rights and give them moral consideration due to human beings," then it would obviously give no guidance unless supplemented with an independent account of rights and due (p. 490) consideration. Some commentators maintain that the formula offers nothing substantial beyond the formula of universal law.[39] Others take the humanity formula to be Kant's basic action-guiding principle, the first principle of morals from which all duties are derivable.[40] Most commentators, however, take an intermediate position, granting that the humanity formula at least adds significantly to our understanding of Kant's basic criteria or procedures for deciding what is right.[41]

Some accounts focus on whether our *purposes can be shared* by those affected by our acts;[42] but others take the key to be a relatively substantive idea of *human dignity;* [43] and still others construe the formula as primarily an imperative to *respect persons*.[44] The apparent differences here are no doubt to some extent matters of emphasis, but some not entirely.

More specifically, the first line of interpretation draws heavily from Kant's claim that we must treat ends-in-themselves as "beings who must themselves be able to share in" our end in acting (Kant, 2002, p. 230 [4:430]). The point, surely, is not that we may do something to others only if they actually share our purpose in doing it, for (as Kant suggests) on this basis a criminal could object to the just sentence a judge imposes. The idea, instead, may be that, before acting in ways that seriously affect others, we must look at our proposed act and purpose from the point of view of a reasonable recipient. We should not act as we propose if the recipient of our act *could not* will for us so to act, that is, could not endorse our so acting without contravening (presupposed) appropriate rational standards for endorsing others' treatment of oneself. On this reading, both the universal law formulas and the humanity formulas ask us, in effect, to assess whether a person could (reasonably, by some appropriate standard) endorse a maxim as a general policy. The universal law formulas focus attention on the perspective of a reasonable agent; the humanity formula directs us to the perspective of a reasonable recipient; but neither sort of reflection is simply about our actual preferences, as agents or recipients.[45]

The second type of interpretation treats the main point of the humanity formula to be an attribution of a special value (*dignity*) to *humanity*, or *rational nature* in human beings. When dignity is treated as a substantive value, this means placing a high priority (above "price") on the preservation, development, exercise, and honoring of our rational capacities.[46] One radical version of this idea treats dignity simply as a high-priority value in a "Kantian consequentialism" devoted to maximizing value.[47] More traditional readings treat dignity as a status of inviolability, not a value that can be quantified and weighed, but a worth to be respected, esteemed, and honored in all our actions. In one version, discussed hereafter, the reasons for acting based on the idea of human dignity are *expressive reasons*, concerned with what our acts *say* and *mean*.[48]

In *The Theory of Morality*, Alan Donagan develops the idea that the humanity formula is a comprehensive principle of respect for persons. Donagan dismisses (p. 491) Kant's universal law formula as a nonsubstantive requirement of impartiality: That is, for a moral system legitimately to treat something as permissible for some persons and not others, there must be "a reasonable ground" in differences between the two groups or their circumstances (1977, pp. 58–59). The highest moral principle, he maintains, is his version of the humanity formula: "[I]t is impermissible not to respect every human being, oneself or any other, as a rational creature" (p. 66). Although granting that judgment is needed to determine specifically what such respect requires, Donagan argues we have enough common understanding of the phrase "respect … as rational" that, without appeal to independent moral principles, we can see that the formula condemns some acts and requires others. In fact, Donagan develops a system of fairly specific moral principles that, he claims, can be derived from the fundamental principle of respect. These draw from Kant's *Metaphysics of Morals*, but modify it considerably.[49] The principles express quite stringent requirements regarding promise-keeping, truth-telling, suicide, murder, use of force, development of talents, beneficence, and various institutional obligations. These principles contain qualifications within them but are not merely prima facie duties. To treat human beings with due respect, we must strictly comply with all the principles.

Even to promote a very good end, we must never use means that violate the principles. We may, however, find ourselves in a moral dilemma because of previous wrongdoing, and then, though we will be acting wrongly no matter what we do, we must do the lesser of evils (pp. 143–149; 1984, pp. 291–309; 1993, pp. 7–21). The humanity formula, Donagan argues, sharply opposes the consequentialist doctrine of negative responsibility: that is, the idea that we are just as responsible for not preventing bad outcomes as for bringing them about directly through our actions. In fact, regarding Bernard Williams's famous story, Donagan thought it absolutely wrong for Jim to kill one innocent Indian to prevent Pedro from killing nineteen others.[50]

Alan Wood also understands the humanity formula as a principle of respect, though Wood holds that all of Kant's formulas should be taken together as a system. He agrees with Donagan that the humanity formula serves a crucial guide for moral judgment and that it is superior, in several ways, to the universal law formulas.[51] Like Donagan, Wood also thinks that some specific act-guiding principles can be derived deductively from the humanity formula when it is understood as a principle of respect for human dignity. We need intermediate premises specifying what it is to respect dignity in different contexts, but he argues that these are empirical and hermeneutical premises, not independent moral principles (1999, p. 154). Since the basic requirement is always to respect humanity in persons, we need intermediate premises specifying the *meaning* expressed by acts in various circumstances. The premises needed in Kant's arguments against suicide and false promises, for example, say that, for various reasons, suicide expresses disrespect for humanity in oneself and making false promises expresses disrespect for another person. Typically, we express disrespect, not by altogether ignoring (p. 492) the dignity of other persons, but by considering our own worth superior to others' or by valuing things of mere *price* above human *dignity*. In Wood's view, there are no algorithms or decision procedures for interpreting the expressive meaning of acts, but we can give reasons for our judgments about this. Kant's humanity formula was not intended to resolve all moral problems directly but to provide a "correct framework" for deliberating about and discussing moral problems (pp. 154–155).

2.3. Illustration Regarding Helping Others

Donagan's treatment of beneficence is typical of his Kantian theory. If we respect persons as rational creatures, we will "take satisfaction in their achieving the well-being they seek, and will further their efforts" as far as we prudently can. So "it is impermissible not to promote the well-being of others by actions in themselves permissible, inasmuch as we can do so without proportionate inconvenience" (Donagan, 1977, p. 85). The ground is not that they deserve it, but that to refuse would be to fail to respect for them as rational creatures. Proper respect, in Donagan's view, requires us to promote the well being or "human flourishing" of others,[52] especially those who, for various reasons, cannot help themselves.[53] Two qualifications are important: it is wrong to promote others' well-being by means that are disrespectful of any rational person, and to promote others' well-being by sacrificing equal or greater goods for ourselves is not required but supererogatory (p. 86). Beneficence, based on respect, also requires "preventing what might harm others or frustrate their permissible projects, and abstaining from actions that would foreseeably elicit responses by which others would be injured" (p. 85). In general, we are required to make it our policy to promote permissible ends of others when there is no disproportionate inconvenience; but in those cases where only we can help another in serious need, without disproportionate sacrifice, then helping that person is morally mandatory.

Wood's treatment of beneficence is similar in some respects, but emphasizes the *meaning expressed* by helping or not helping others. In his view, "the reason that we should help others in need is that we thereby *exhibit* proper esteem for their worth as rational beings" (1999, p. 149, emphasis added). Because of this, he argues, "rugged individualists" who want no one to be charitable cannot rationally will this as a universal policy, because this would entail willing that others "show contempt" for their humanity if they should fall into dire need (p. 150). Presumably, not helping others in many circumstances would not show contempt for them, and so the duty to aid others is limited in various ways. Only when not helping is disrespectful of their humanity is it morally mandatory, and this is a matter of interpretation.

(p. 493) 2.4. Problems and Doubts

Disputes about which interpretations best explain Kant's texts may be endless. For present purposes, I want merely to call attention a few potential problems in different ways of developing and relying on Kant's humanity formula.

First, and perhaps most obviously, the formula presents an ideal that is important and inspiring but far from a determinate, precise decision procedure. Virtually all commentators acknowledge this to some extent, but some (for example, Donagan) may have more confidence than warranted in our ability to agree on what specifically "respecting a person as rational" implies for a wide range of cases. This becomes more of

a problem the more the humanity formula is taken to be a self-standing guide to decision in all particular cases or by itself a fully determinate first premise for justifying strict moral rules.

Second, if we treat the humanity formula, as Donagan and others do, as primarily about relatively simple, self-contained interpersonal exchanges, we risk missing relevant moral considerations that arise from the larger context. What seems on the surface to be a disrespectful way to treat one person (e.g., a lie) may have ramifications regarding others that could justify it from a perspective that is concerned with how best to express equal respect for all persons. From that perspective, even the person deceived might approve the general policy that would allow exceptions to the duty of truth-telling for that occasion. Exclusive focus on what is intuitively respectful in the narrower context of person-to-person interactions leads readily to apparent dilemmas in which, so far as we can tell, all of our options are forbidden.[54] Arguably, however, we can make progress on some of these problems if we treat the humanity formula as an evaluative attitude mandated first and foremost for our deliberations about general moral principles and policies in the light of the many complex factors that may be relevant.[55]

Third, the more the humanity formula is understood to express substantive, "thick," and controversial values, the harder will be the task of arguing convincingly that everyone, despite cultural and individual differences, has good reasons to accept it. Sometimes, of course, it is hard to convince everyone of propositions that are nevertheless true and backed by good reasons, and so a strong burden of proof is not the same as disproof. Nevertheless, arguably, both Donagan's stringent nonconsequentialist "Kantian" rules and Cummiskey's "Kantian" consequentialism present the humanity formula as essentially committed to (opposite) sides of some issues about which reasonable persons disagree. For example, even with its (Kant-inspired) value priorities, Cummiskey's Kantian consequentialism seems open to the familiar objection that there are conflicts between justice and what promotes the best consequences. Similarly, Donagan's (relatively) inflexible principles apparently would require, in certain emergency situations, refusal to protect lives and other Kantian values by extraordinary means that reasonable persons might rightly (but reluctantly) approve. Donagan more nearly captures Kant's own (p. 494) particular moral beliefs, but these are not necessarily implicit in the humanity formula or correctly derivable from it. Both substantive theories have the burden of defending *why* some apparently reasonable common moral judgments are wrong.

Finally, apart from this, there is philosophical reason, as well as textual evidence, for treating Kant's supreme moral principle (in all its versions) as describing an essential framework for moral deliberation and discussion rather than as an independent and determinate guide, as some have assumed.[56] In a world of tragic cultural and individual conflicts, moral philosophy needs to articulate a point of view from which, despite substantial differences, human beings can work together toward reasonable, mutually acceptable principles and policies. This purpose is defeated if a theory *too readily* renders judgment on a wide range of controversial moral issues. This is not to say that a moral

theory should not in the end sharply condemn certain conventional moralities. The point is that moral theorists should be careful not to unduly allow their own strong moral convictions about particular issues to shape their ideas about what is essential to a moral point of view. If, as Kant thought, morality concerns what reasonable persons can accept, despite diversities, then accounts of the basics of morality should leave room for diversity and reasonable disagreement regarding many particular situations.

3. A Kantian Legislative Perspective

3.1. The Formulas of Autonomy and a Kingdom of Ends

Kant's next two formulas are closely related. In fact, Kant did not even distinguish them in his review of formulas. The *first*, commonly called the formula of autonomy, is initially characterized as "the Idea of *the will of every rational being as a will that legislates universal law*" (Kant, 2002, p. 232 [4:431]). The point, evidently, is that we must *always act in ways compatible* with this idea. The context makes clear that Kant conceives the rational law giving in question as not motivated by inclinations and desires. There are some differences of interpretation among commentators, especially as the abstract formula of autonomy is reexpressed through the somewhat fuller idea of a kingdom of ends (pp. 233–234 [4:433]). The following points, however, seem fairly clear.

The independence of desires and inclinations that the formula attributes to (p. 495) rational lawmakers is not a total lack of motivating dispositions. It would be absurd to think of them as legislating while "caring" about nothing. All rational agents have ends; this is inherent in the very idea of a rational agent. In addition to whatever (rationally) contingent ends they may have as individuals, rational agents as such are (allegedly) disposed to acknowledge a rationally necessary ("objective") end-in-itself, which is "humanity" or "rational nature." In Kant's view, unlike Hume's, reason is not "inert." Kant holds that "pure reason can be practical": That is, rational agents as such, independently of contingent inclinations, are supposed to be necessarily committed to basic noninstrumental, action-guiding principles, expressed in the formulas of the Categorical Imperative. The values inherent in this commitment can be a motivating factor, independently of our desire-based ends.[57]

In addition, although Kant's claim that there can be rational law-giving not motivated by inclinations is controversial, it does not imply the process takes place in complete *ignorance* of human inclinations. Rational agents with autonomy are lawmakers in some respects like ideal secular or divine legislators, though their "laws" are not legal requirements backed by threats of punishment but rather moral principles endorsed for themselves (as well as for others). Rational legislators of universal laws do not choose their laws as a means to satisfying their own special inclinations, but, at some stage, they must take into account the conditions under which the laws will be applied. All of the formulas of the Categorical Imperative are supposed to express basic objective principles the rationality of which does not depend on contingent human conditions, but none can be applied to particular human situations without some knowledge of that situation and general facts about our world. Many particular facts may prove to be irrelevant to a given judgment, but the formulas cannot guide us to any particular judgment until we consider the context of application.

Rational lawmakers are supposed to give themselves laws as rational wills with *autonomy*. This implies that they are committed to standards of rational choice beyond hypothetical imperatives. They do not merely acknowledge the rationality of taking necessary means to their desired ends; they have over-riding rational standards of the

form "One ought to do X, and not just because it promotes one's desire-based ends." They have the capacity and predisposition to act on principles that are not based on their desires as individuals. Further, all such principles are ultimately grounded in rational requirements (expressed in the Categorical Imperative) that are independent of all rationally contingent features of human nature. Human moral agents, being imperfectly rational, do not always live up to these principles that express their autonomy of will, but inevitably, Kant thought, they at least implicitly acknowledge the principles as standards that should override any conflicting claims of inclination. In extremely immoral persons, this acknowledgement may be more evident in their pangs of conscience (p. 496) than in any resolution to reform; but if some people lack it completely, this (in Kant's view) would indicate that either they lack the capacity for practical reason or (like very young children) they have not developed it.[58]

The idea of autonomy implies further that the legislators' laws are, in a sense, *self-imposed*. We need to distinguish here basic principles that are constitutive of rational agency from the more specific "laws" that the rational legislators are supposed to give to themselves. The former presumably include the other forms of the Categorical Imperative, the hypothetical imperative, and any other principles necessary to rationally consistent and coherent choice. The latter are the more specific moral principles that, given appropriate facts, rational legislators (who are committed to those basic standards) would adopt for specified (e.g., human) conditions. (The *Groundwork* is primarily devoted to the basic principles, and *The Metaphysics of Morals* is primarily devoted to intermediate level principles or "laws" for human conditions.) We cannot coherently think of the rational legislators as "making" and "giving themselves" the basic rational standards that constitute their rationality. Nevertheless, their autonomy implies that they *identify* themselves with these constitutive standards rather than seeing them as externally imposed and alien. When their reasoning according to the standards leads to a particular judgment, then they regard it as the result of their own judgment. The more specific "laws" that the legislators make can be understood as *given to themselves by themselves* in a further sense that is represented in the idea of a kingdom of ends. Here for good reasons, all rational legislators endorse the same general laws for everyone, and they are subject only to laws that they all endorse as legislators. In a sense, then, they legislate for themselves rather seeing laws as given to them by nature, tradition, or divine command.[59]

Kant does not give examples to show how the formula of autonomy might be applied. Instead, he develops its core idea into a *second* formula expanding the idea of autonomy, the kingdom-of-ends formula. This appears in his review of the formulas as a "complete determination of all maxims": "*All maxims which stem from autonomous lawgiving are to harmonize with a possible kingdom of ends as with a kingdom of nature*" (Kant, 2002, p. 237 [4:436]). A kingdom of ends is "a systematic union of different rational beings under common laws" (p. 234 [4:433]). It is an ideal analogue of a political state or commonwealth in which rational members make all of the laws to which they are subject. The members have "private ends," but in conceiving of the kingdom of ends we "abstract" from their content as well as from other "personal differences" among the members (p.

234 [4:433]). The kingdom has a nominal sovereign or head that makes the same laws without being subject to them. Because the sovereign (a "holy will") lacks needs and inclinations that might interfere with rational choice, the laws of the kingdom do not appear as *imperatives* or *obligations* for the sovereign (p. 240 [4:439]).[60] Other (human) members are subject to obligation because, though subject to contrary inclinations, they too are committed to the same rational (p. 497) principles. The sovereign is conceived as having "unlimited resources adequate to his will." It may help to motivate us, Kant thought, to think of the kingdom of ends and the kingdom of nature as united under this same sovereign (pp. 234–235 [4:434], pp. 239–240 [4:439]). The kingdom of ends, Kant says, would become actual if everyone always followed its laws, but we are still obligated to follow the laws of a possible kingdom of ends in our world, where obviously not everyone does so (pp. 238–239 [4:438]).

3.2. Interpretative Issues

There are a number of questions about how to interpret these formulas. For example, how are they related to the preceding formulas? In his review of formulas, Kant suggests that the universal law formula is the best to use in judging maxims. The later formulas help to provide "access" to the moral law by bringing it "closer to intuition" and "feeling" (Kant, 2002, pp. 236–238, [4:436–437]). He also suggests, however, that the kingdom–of–ends formula combines ideas in the preceding ones and so provides the most "complete determination" of moral requirements on maxims (p. 237 [4:436]).[61] The former suggests that the main function of the kingdom–of–ends formula is to inspire and motivate us, but the latter suggests that it gives Kant's fullest characterization of the moral requirements on maxims and so, presumably, our most comprehensive action-guide.

Again, are we to think of all moral agents as members of the kingdom, or only those who show themselves worthy of it by acting morally? At one point, Kant suggests that the possible kingdom would become actual only if we all did our duty; but later he suggests that it would become a "true reality" only if God were sovereign of both nature and the kingdom of ends, thereby ensuring that the permissible ends of the virtuous would be fulfilled (Kant, 2002, pp. 238–240 [4:438–439]). For practical purposes, however, what matters is *what we as rational members would legislate*, not whether we think of ourselves and others as actually or merely possibly belonging to the kingdom of ends.

The most important interpretative question, especially for later development of Kantian ethics, concerns how the members in the kingdom make their laws. Kant's political metaphors suggest an ideal legislature in which all citizens are free and rational legislators and (except for the sovereign) bound by the laws that they make. As legislators, they share the same rational dispositions and values (e.g., humanity as an end-in-itself). Since the model abstracts from their personal differences and the content of their private ends, the laws that one legislator would endorse are just the same as the laws that each other would endorse. Nonetheless, the picture is that of joint legislation by and for all members. Their agreement results, as it were, from the fact that they

acknowledge the same reasons and (p. 498) cannot be sidetracked by private concerns not shared or recognized by others. Here "abstracting" from differences functions, rather like Rawls's "veil of ignorance," to guarantee impartiality at the highest level of moral deliberation (1971, pp. 11, 17, 188–223). On this reading, the kingdom-of-ends formula would tell us always to conform to those laws that we would make as members of this ideal moral legislature. The legislature would differ in significant ways from real legislatures. For example, the legislators do not make coercive laws backed by sanctions; their jurisdiction includes (at least initially) all rational agents;[62] the legislators are not bound by independent obligations or subject to external pressures; they never legislate irrationally; and so on.

Some passages, however, suggest a different picture. Here we imagine a possible world in which everyone conforms to the universal law formula, acting only on maxims that they can will as universal law. The outcome, we are to suppose, would be a harmonious system of ends, with no one being treated as a mere means and everyone having a chance to pursue their permissible ends without undue interference. Since there is no legislature in this picture, arguably there is no need to ignore personal differences to ensure impartiality. The "laws" of the kingdom would be identified as individuals tested their maxims and found that certain ones could not be willed as universal law. "Never kill people for profit," for example, would be a law if no one could will a proposed maxim of killing for profit as a universal law. This interpretation allows the universal law formula to do the work of moral assessment, leaving for the kingdom-of-ends formula the role of an inspiring idea of the outcome if we all would do our part.[63]

3.3. A Legislative Model for Deliberation about Moral Rules

On close examination, the texts remain ambiguous, and they leave many questions open. On the last interpretation, however, the kingdom of ends offers no new practical guide to deliberation and it inherits all of the problems of applying the universal law formulas. Arguably, however, by construing the kingdom of ends instead as an idealized model of moral legislation and then extending this idea, we can take Kantian ethics (broadly conceived) in a different, and perhaps more promising direction. The inspiration for the project comes from John Rawls's theory of justice, although Rawls himself interprets Kant rather differently.[64]

The project presupposes that we can distinguish fundamental Kantian values from less basic rules and principles concerned with particular areas of the moral life. From the former, drawn primarily from versions of the Categorical Imperative, we would try to construct a conception of the perspective from which we should reflect about more specific rules and principles, how to articulate them, what qualifications they should contain, and so on. We would think of more basic (p. 499) values as constituting, for purposes of the theory, the essentials of an ideal point of view for trying to work out with others what moral demands and limits should be included in our principles concerned, for example, with obedience to the law, property rights, just punishment, gratitude,

beneficence, respect, and friendship. It should not be assumed, of course, that we need specific principles or rules for all areas of life. Working toward common, well-defined standards is important in some areas; but in others, we may do better simply to approach problems with a good attitude and sensitivity to the context. In addition, the principles, if any, that we can justify as reasonable independently of historical context may be few and indeterminate in ways that leave much room for interpretation and judgment. In any case, despite what Kant himself thought, the best we can do may be to use the basic deliberative perspective to work on problems that arise in the more circumscribed conditions in which we now live.[65] Reasonable principles for familiar circumstances may be applicable more generally, but extending them is impossible unless we understand the other contexts well enough to see that there are no relevant differences.

Although we might expect that actual people who take up the deliberative perspective could agree on some general principles that rather directly reflect their constitutive values, we cannot expect that they would always agree. Agreement on fundamental values does not ensure unanimity on how these can be best implemented in various situations. Differences in judgment and in understanding of background facts are bound to result in disagreements on particular moral issues. The legislative perspective may still be useful, however, as a standard for conscientious action. That is, we can think of conscientious action as doing whatever, after due deliberation and dialogue with others, we judge should be required by principles adopted from the legislative perspective. We would need to consider what we would recommend for universal approval if we were debating the matter from the legislative perspective. Then, as conscientious persons, we should follow this, keeping in mind that others may reasonably disagree.

The Kantian legislative perspective suggested by the kingdom-of-ends formula combines ideas from Kant's other formulas. *First*, as legislators, we deliberate by assessing whether we can reasonably endorse various general policies, considered as principles for everyone. Thus we must move beyond the narrower point of view that we take when first inclined to act in a certain way and consider how things would be if we endorsed moral principles permitting everyone to act that way. In doing, so we need not identify precisely "the maxim" of each act, but instead can review systematically sets of permissive, prohibitive, and obligatory principles together. Although our maxim might be described in any of several ways, our act will be wrong if, under any description, it is incompatible with the principles that we would endorse from the legislative perspective.

Second, legislators who acknowledge each other as ends-in-themselves would have to deliberate with the constraints and values, whatever these may be, inherent (p. 500) in the idea of humanity as an end-in-itself. Because the legislators have "private ends" but "abstract from differences," they must generally favor conditions that further the members' ends, but their law-making should not be influenced by partiality for their own ends in particular. In addition, arguably, the value of humanity as an end motivates them to place a high priority on members' survival, development, and opportunities to live as rational agents. More abstractly, the value of humanity is reflected in their unwillingness to engage in conduct that they could not, from the legislative perspective, justify to other

members. This motivating disposition to treat humanity in each person as an end would be considered not merely a contingent desire but rather a disposition inherent in rational nature.

Third, the idea of autonomy is reflected in the stipulation that the legislators are subject to no laws or principles other than those they themselves rationally endorse. They do not endorse principles merely because tradition, the state, or God sanctions them. They endorse some fundamental rational constraints and values, such as humanity as an end, simply because these express their nature as rational persons. They endorse more specific moral principles, such as beneficence and fidelity, because these principles prescribe conduct that expresses, secures, or promotes their basic values in the context of our complex empirical world. The idea of autonomy is also reflected in the stipulation that, when deliberating, legislators set aside inappropriate partiality toward their own special ends. Further, valuing other persons as (at least potentially) autonomous colegislators of the moral standards that govern our mutual relations has implications regarding specifically how we are to respect them as persons.

3.4. Illustration: Meeting Needs and Doing Favors

The Kantian legislative perspective sketched here is obviously not yet an adequate normative ethical theory, much less a complete one. Let us, however, overlook for now remaining gaps and problems and return, instead, to the example of beneficence. From a Kantian legislative perspective, how might we think about grounds and limits of our obligation to help others? Let us consider what general duty we should attribute to everyone, acknowledging that special circumstances are likely to call for more or less than any general principles can anticipate.[66]

Consider several candidates regarding the general obligation to help others in need. (1) We should always help those in need when we can, whether they have a right to it or not. (2) Helping others in need is morally optional, except when they have a right to demand it. (3) We should help others with basic needs, at least when their need is great and the cost to us is proportionately small. As legislators concerned to work out a reasonable system of moral principles, we (p. 501) might at once see that all of the principles need to be qualified by adding (a) *except that helping is generally wrong when their need is for resources to complete immoral projects and* (b) *except that it is wrong when we can only help by immoral means*. "Immoral" here would have to be determined by other, higher priority principles that the legislators have adopted. On further reflection, we might see that what others need to survive, develop, and thrive as rational autonomous agents (i.e., "basic needs") should take priority over what they need to accomplish projects based only on personal preference. This seems implicit in valuing rational nature as an end-in-itself, at least on substantive interpretations. Especially within the first category, it seems that the greater the need, other things equal, the more reason to help. Given this, we might realize that the same priority applies equally well to the agent's own needs. Thus, we should not in general disregard our urgent basic needs merely to help others in minor,

optional projects. Thus the first candidate (1), even amended with (a) and (b), would be unacceptable, because it ignores the agent's own needs and potential differences in the kind and degree of need. If amended to take these factors into account, (1) would look more like (3).

We cannot evaluate (2) properly without first deciding what help persons can demand as their right. This would have to be a prior issue to take up from the legislative perspective, but let us assume for now that familiar intuitions about this are right. In general, since rights are enforceable claims, presumably moral legislators would want to express and promote their values not only by assigning individual rights but also by adopting principles that call upon conscientious persons to do more for others than they can demand by right.[67] If so, candidate principle (2) would be unacceptable. It implies that, beyond doing what we can rightfully be coerced to do, we may do as we please, ignoring all the needs of others, no matter how great these are and no matter how little it would cost us to help. If everyone, even everyone so inclined, were to act that way, then the prospects of everyone to live a full life as a rational, autonomous person would be diminished. Secure, wealthy individuals might calculate that they are likely to be better off under (2) than (3), but they could not justify that idea to others, or themselves, if they took up the legislative point of view that abstracts from personal differences and views humanity in each person as an end.

The third candidate (3), by contrast, seems quite promising. If this principle (at least) is endorsed, then, given the uncertainties of human life, the prospects of every representative person are better in ways that matter to rational legislators. The main question about (3) is whether it is too minimal a requirement, and so needs to be supplemented with further principles regarding helping.

Two supplements, at least, seem to be appropriate from the legislative perspective. First, because (3) addresses only cases where we could help others whose needs are great and basic to rational, autonomous living, we need some general principle about helping others whose needs are not as urgent and fundamental. We need to consider helping others in their personal projects: for example, doing (p. 502) favors rather than attempting to meet basic needs. As before, legislators will want to qualify any principle in this area by ruling out helping others in their immoral projects or by using immoral means, where immorality is defined by incompatibility with higher priority principles, already endorsed. Consider three candidate principles concerning helping others where great basic needs are not involved: (4) Help others whenever you can, unless fulfilling some other duty. (5) Helping others is entirely optional, assuming that you fulfill your other duties. (6) Make it an end of yours that others realize their (permissible) ends; and so, to some significant extent, contribute to others' (permissible) ends as well as your own.

A problem with (4) is that it fails to acknowledge the importance, from the legislative perspective, of the agent's own personal ends. To be sure, (4) implies that others should help the helper, and so the helper's ends are not totally ignored. However, (4) also unreasonably implies that, other duties aside, each person should help to promote others'

ends, irrespective of the cost to the person's own projects. Thus, everyone would have to sacrifice their very important projects whenever they could permissibly help someone else on a relatively minor project. The principle must at least be adjusted to require only sacrificing opportunities to promote one's own projects when, all considered, helping others instead would contribute proportionately more to others' equally or more highly valued projects. Even amended this way, (4) is still a far more demanding requirement than most people could accept. Unless we are working to satisfy other duties, it requires us to drop whatever we are doing whenever an opportunity presents itself to work on another person's personal project, provided only the other person cares slightly more about his or her project than we care about ours. A readiness to do so might seem saintly, in some respects, but, given the limits of human nature, it seems more than could reasonably be expected of everyone. In addition, being constantly "on call" to help others who are not in urgent or vital need would make it difficult, if not impossible, to structure a life with meaningful personal projects of one's own, which is something that presumably rational autonomous agents would value.

Given that rational agents necessarily set themselves and pursue ends, candidate principle (5) arguably expresses an attitude incompatible with valuing rational agency in each person as an end-in-itself. It implies that we may ignore the (permissible) projects of others, no matter how much the others care about them and no matter how insignificant the inconvenience of helping would be to ourselves. If everyone, or even those so inclined, always acted in this way, everyone's prospects, considered from the legislative perspective, would be worse than if they adopted candidate principle (6). Some fortunately situated individuals might prosper, but legislators respecting humanity in each person equally would not cater to their special advantages. Arguably they would acknowledge the worth of all as rational, autonomous persons by requiring everyone to aid others in personal projects, to some significant extent, but leaving a wide latitude for choice (apart (p. 503) from the requirement regarding basic needs) as to when, how, and exactly how much to help. Principle (6) is just this sort of requirement, for it tells us to include it among our important ends that others also fulfill their (permissible) personal ends.

From the Kantian perspective, persons have a status of "unconditional and incomparable worth" that is acknowledged in practice by conforming to the "laws" to which they all, as rational and autonomous legislators, would agree. They are "equal," in that they have the same status under the moral law, no one being intrinsically more important than any other, and each person being inviolable in ways determined by the principles everyone mutually endorses. This equal standing under the moral law is quite different from the sort of equality that classic utilitarians prescribe. Equality for them is equal consideration (or "weight") for equal pleasures and pains, no matter whose they are. That idea is not part of the Kantian picture, and, as is often noted, following it could lead to exploiting and oppressing a few for the greater pleasure of many. Kantian legislators are not utility-maximizing legislators but are constrained by their acknowledgment of a worth of individual persons that cannot be quantified, weighed, or traded for "more" value. The specific implications of this idea still need to be worked out, but it clearly implies that

Subscriber: King's College London; date: 24 November 2018

Kantian legislators do not have the authority to prescribe *whatever* is predicted to produce a maximum quantity of some comparable and conditional value, such as pleasure, satisfaction, and the like. Further, rejecting the idea that they must measure, compare, and produce the greatest quantity of happiness, Kantian legislators need not even require the qualified principle "You must always help others whenever other duties allow and others would gain more happiness than you would lose by helping." By contrast, the more indeterminate principle (6) expresses respect for all persons by valuing their (permissible) ends without so restrictively dictating precisely when and how much time and energy to devote to helping others, beyond the requirement regarding basic needs (3).

3.5. Remaining Problems

These last comments are meant merely to illustrate how reflection on a Kantian legislative perspective might proceed. Further considerations are no doubt relevant, and in any case the legislative perspective has not been defined fully enough to permit more than rather loose, informal argument. In addition, my discussion has concerned only the general principle of beneficence, not special requirements in specific circumstances. All attempts to reconstruct Kant's basic normative standard must work out remaining problems and address various objections. The proposed idea of a Kantian legislative perspective is obviously no exception. For example, we need to consider further the importance and limits of rules and (p. 504) principles for a moral life. The appropriate degree and kind of "abstracting from personal differences" needs to be clarified, especially as this may vary with the context and scope of the problem to be addressed. Different ways of understanding the central value of humanity as an end-in-itself need to be clarified and assessed. Any other presupposed standards of "rational" and "reasonable" willing need to be articulated explicitly. Special problems are raised when we try to use and adjust principles for different social and economic conditions. If we insist on quite simple, inflexible principles appropriate for an idealized world, we may unreasonably ignore morally relevant differences between those conditions and the real world. If we bend and adjust principles too readily to accommodate special circumstances, there would be no point in having principles. It would be madness to insist that the consequences of adopting a principle never matter, but it must be clear how and within what limits reflection on consequences is appropriate from a Kantian legislative perspective. We also need to address suspicions that trying to construct principles from an ideal perspective complicates moral judgment needlessly and alienates us from our more particular moral perspective in daily life.[68] Much can be said in response to these problems, but whether the problems ultimately undermine the promise of this approach to Kantian normative theory remains to be seen.

References

Baron, Marcia. 1995. *Kantian Ethics (Almost) without Apology.* Ithaca, N.Y.: Cornell University Press.

Brandt, Richard. 1959. *Ethical Theory.* Englewood Cliffs, N.J.: Prentice Hall.

Cummiskey, David. 1996. *Kantian Consequentialism.* New York: Oxford University Press.

Dean, Richard. 1996. "What Should We Treat as an End-in-Itself?" *Pacific Philosophical Quarterly* 77: 266–288.

Donagan, Alan, 1977. *The Theory of Morality.* Chicago: Chicago University Press.

Donagan, Alan. 1984. "Consistency in Rationalist Moral Systems." *Journal of Philosophy* 81: 291–309.

Donagan, Alan. 1993. "Moral Dilemmas, Genuine and Spurious: A Comparative Anatomy." *Ethics* 104: 7–21.

Guyer, Paul, ed. 1998. *Kant's Groundwork of the Metaphysics of Morals: Critical Essays.* Lanham, Md.: Rowman and Littlefield.

(p. 513) Herman, Barbara. 1984. "Mutual Aid and Respect for Persons." *Ethics* 94: 577–602. Reprinted in Herman, 1993, 45–72.

Herman, Barbara. 1993. *The Practice of Moral Judgment.* Cambridge, Mass.: Harvard University Press.

Hill, Thomas E., Jr. 1979. "A Review of *Acting on Principle* by Onora Nell." *Ethics* 89: 306–311.

Hill, Thomas E. 1980. "Humanity as an End in Itself." *Ethics* 91: 84–90. Reprinted in Hill, 1992a, 38–57.

Hill, Thomas E. 1992a. *Dignity and Practical Reason in Kant's Moral Theory.* Ithaca, N.Y.: Cornell University Press.

Hill, Thomas E. 1992b. "A Kantian Perspective on Moral Rules." In *Philosophical Perspectives*, vol. 6, *Ethics*, ed. James E. Tomberlin, 285–304. Atascadero, Calif.: Ridgeview. Reprinted in Hill, 2000, 33–56.

Hill, Thomas E. 1992c. "Making Exceptions without Abandoning the Principle: Or How a Kantian Might Think about Terrorism." In Hill, 1992a, 196–225.

Hill, Thomas E. 1993. "Donagan's Kant." *Ethics* 104: 22–52. Reprinted in Hill, 2000, 119–152.

Hill, Thomas E. 1997. "A Kantian Perspective on Political Violence." *Journal of Ethics* 1: 105–140. Reprinted in Hill, 2000, 200–236.

Hill, Thomas E. 1999. "Happiness and Human Flourishing in Kant's Ethics." *Social Philosophy and Policy* 16: 143–175. Reprinted in Hill, 2002, 164–200.

Hill, Thomas E. 2000. *Respect, Pluralism, and Justice: Kantian Perspectives.* Oxford: Oxford University Press.

Hill, Thomas E. 2001. "Hypothetical Consent in Kantian Constructivism." *Social Philosophy and Policy* 18: 300–329. Reprinted in Hill, 2002, 61–98.

Hill, Thomas E. 2002. *Human Welfare and Moral Worth: Kantian Perspectives.* Oxford: Oxford University Press.

Kant, Immanuel. [1785]. 1964. *Groundwork of the Metaphysics of Morals.* Trans. and ed. H. J. Paton. New York: Harper and Row.

Kant, Immanuel. [1797–98]. 1991a. *The Metaphysics of Morals.* Trans. and ed. Mary Gregor. Cambridge: Cambridge University Press.

Kant, Immanuel. [1793–94]. 1991b. *Religion within the Boundaries of Mere Reason.* Trans. and ed. Allen Wood. Cambridge: Cambridge University Press.

Kant, Immanuel. [1788]. 1997. *Critique of Practical Reason.* Trans. and ed. Mary Gregor. Cambridge: Cambridge University Press.

Kant, Immanuel. [1785]. 2002. *Grounding for the Metaphysics of Morals.* Trans. and ed. Thomas E. Hill, Jr., and Arnulf Zweig. Oxford: Oxford University Press.

Korsgaard, Christine. 1996a. *Creating the Kingdom of Ends.* Cambridge: Cambridge University Press.

Korsgaard, Christine. 1996b. "Kant's Formula of Universal Law." In Korsgaard, 1996a, 77–105. Originally published in *Pacific Philosophical Quarterly* 66: 24–47.

Nell, Onora (O'Neill). 1975. *Acting on Principle.* New York: Columbia University Press.

O'Neill, Onora. 1989. *Constructions of Reason.* Cambridge: Cambridge University Press.

Paton, H. J. 1958. *The Categorical Imperative.* London: Hutchison.

Pogge, Thomas. 1998. "The Categorical Imperative." In Guyer, 1998, 189–214.

Rawls, John. 1971. *A Theory of Justice.* Cambridge, Mass.: Harvard University Press.

Ross, W. D. 1930. *The Right and the Good.* Oxford: Clarendon Press.

Singer, Marcus G. 1961. *Generalization in Ethics.* New York: Knopf.

Kantian Normative Ethics

(p. 514) Williams, Bernard. 1973. "A Critique of Utilitarianism." In *Utilitarianism: For and Against*, ed. J.J.C. Smart and Bernard Williams, 77–155. New York: Cambridge University Press.

Wolf, Robert Paul. 1973. *The Autonomy of Reason.* New York: Harper and Row.

Wood, Allen W. 1999. *Kant's Ethical Thought.* Cambridge: Cambridge University Press.

Notes:

(1.) For example, are moral judgments based on reason or sentiment? What does it mean to say that something *ought* to be done? What distinguishes moral from nonmoral "ought" claims? Do moral predicates refer to properties that are objective and real? Are they empirically accessible natural properties or constructs of some kind? The terms we use in our contemporary discussions are often different from Kant's, but Kant clearly had positions and arguments that are relevant to these issues. Some may doubt that metaethical and normative questions can be separated and answered independently, but I will proceed as if they can be usefully addressed, at least to some extent, as separate matters. My project, then, is to survey Kantian answers to the normative questions, leaving aside for now questions about how these are related to other issues.

(2.) My discussion of beneficence here, though inevitably sketchy, may help in a small way to fill a gap in my previous accounts of the duty of beneficence in Kantian ethics. See my *Human Welfare and Moral Worth: Kantian Perspectives*, 2002, chs. 3, 5, 7. The last of these essays, "Meeting Needs and Doing Favors," addresses the recent controversy over whether this duty to promote others' happiness should be interpreted as a very stringent ("rigoristic" or "robust") requirement or as a quite limited ("minimal" or "anemic") requirement. Following H. J. Paton and Mary Gregor, I argued earlier that the duty to promote the happiness of others, at least as presented in Kant's fullest account of it in *The Metaphysics of Morals*, 1991a, is a minimal requirement. David Cummiskey and, to some extent, Marcia Baron have argued that the requirement is morestringent. My contention is that, unsurprisingly, we find a modest beneficence requirement in *The Metaphysics of Morals*, for that work is supposed to describe duties in the most general way appropriate for all human conditions. A modest general requirement makes good moral sense, however, only if there is a plausible way to argue from basic principles that in various special circumstances helping others is not optional. By using beneficence as my example when I survey applications of Kant's basic moral principle(s), I can show various ways in which Kantians can argue from the Categorical Imperative to this conclusion. These arguments may help to fill the gap in my previous defense of the minimal interpretation of the general principle of beneficence, for, if successful, they would supplement Kant's minimal general principle with Kantian grounds for judging that in many cases, giving aid to others is strictly required. See Baron, 1995, and Cummiskey, 1996.

(3.) Numbers in brackets refer to the corresponding volume and page numbers in the standard Prussian Academy edition.

(4.) Wood, 1999, pp. 363–365 plausibly suggests that some of Kant's statements that are commonly taken to be versions of FUL are in fact expressions of his later formula of autonomy.

(5.) Later Kant restricts the term to the "material principles" that agents act on, i.e., principles based on their (nonrational) desires. See Kant, *Critique of Practical Reason* (hereafter abbreviated CPrR), 1997, pp. 17–19 (5:19–22).

(6.) Kant's own examples are not uniform in structure. They usually describe the act and indicate its purpose but only sometimes state a separate reason or motive. In Kant's suicide example, 2002, p. 223 (4:422), the act is suicide, the purpose to end pain (when life promises no more compensating pleasures), and the underlying reason or motive is "self-love."

(7.) O'Neill, 1989, p. 84, makes a case that the maxims that should be tested are "underlying principles or intentions by which we guide and control our more specific intentions." This suggestion that they express our more general, fundamental life commitments is in line with Kant's suggestion in CPrR that all nonmoral maxims fall under "the general principle of self-love or one's happiness" (O'Neill, 1989, p. 84; Kant, 1997, p. 19 [5:22]). Kant himself does, however, use the universal law formulas to test examples that are more specific.

(8.) Here I set aside controversies about whether FUL proposes a test for maxims different from that proposed in FULN.

(9.) The reason for the first point is that, for purposes of her deliberation, the maxim must reflect the agent's understanding of what she is doing (and why). (Kant often suggests that we cannot know with certainty what maxims we were really acting on in the past, but in deliberating about what to do next we must assume that we will act on whatever maxim that we adopt after due reflection.) The second point stems from the fact the very purpose of trying to articulate our maxims is to determine for ourselves whether it is morally all right to act as we are inclined to. It would defeat our purpose to work with a phony maxim that not even we ourselves see as accurately reflecting what we intend to do, our reasons, and factors we count as morally relevant. Suppose you plan to gain political power to serve your ethnic community but you know that taking the necessary means will harm many innocent people. A maxim that omitted reference to the fact that your act would harm innocent people would not be the morally appropriate one to test. (This fact should turn out to be salient anyway,when you try to conceive and will the world where everyone acts on your *unqualified* maxim to take the available means to help your ethnic community.) If the third point were ignored, the reasons we could will certain maxims as universal law might have nothing to do with their moral status. For example, one might (mistakenly) argue that because irrelevant details make unwelcome recurrences of the case unlikely, you might will as universal law the maxim "To convenience myself, I shall tell a lie to a bald, skinny, diabetic man on a Tuesday night with a full moon." If we dishonestly include morally irrelevant detail or omit salient moral considerations, the universal law tests are likely to give morally inappropriate results. However, the apparent need to make judgments of moral relevance before applying the universal law tests poses a problem for those who think that the tests are sufficient by themselves to determine right and wrong. Herman, 1993, addresses this problem at length.

(10.) The sort of limits I have in mind may be expressed, for example, in the maxim "I will refuse a government order to do something *if what was ordered was contrary to an already established perfect ethical duty* (such as to bear false witness in court)." This is a qualification or limit that Kant acknowledged (1991a, p. 98 [6:322], pp. 136–137 [6:371]; 1998, p. 153 n. [6:154 n.]). Despite what Kant himself thought, it seems only reasonable that the maxim of person who wants to help an innocent friend escape from a murderer might be "I will tell a lie *if it is told to someone who is threatening the life of an innocent person and if it is the only way, as far as I can tell, to prevent a murder.*"

(11.) That is, cannot be conceived as either a universal law or as a universal law of nature (if that is different). For present purposes, I am setting aside the question whether the formulas offer two tests or only one.

(12.) Kant suggests that failure to pass the first (contradiction in conception) test indicates a violation of strict or "perfect" duty. Maxims that fail the second (contradiction in willing) test are bad maxims that we should not act on, but the general principles that we would violate by doing so are principles of "imperfect" duty. These require us, for example, to adopt the maxims to promote as ends the happiness of others and our own

perfection. It is strictly a duty to adopt these maxims, but what we must do toward the prescribed ends is not specified in a determinate way.

(13.) The maxim as stated here is perhaps too simple and unqualified to be reasonably willed as a law for all circumstances. We may, for example, have conflicting obligations; others may be better positioned to provide the aid; etc. For now, I set aside such complications. For any explicit maxim, there will presumably be background conditions implicitly understood. Any attempt to defend the universal law formulas as satisfactory determinants of right and wrong would have to confront the problems these implicit background assumptions raise.

(14.) Suppose, instead, in considering not aiding accident victims (even though only I am available to help them), I describe my proposed maxim merely as "To save time, I will not make stops on my drives to work." Because this omits the morally salient fact that not stopping now may result in the death of innocent persons, the maxim may seem quite innocent, that is, one that we can conceive and will as universal law. Nevertheless, it seems clear that it would be wrong not to aid the injured persons. (Arguably, if the maxim I stated is intended as an inflexible policy, we should be able to see that it cannot be reasonably willed as a universal law, despite its innocent look; for, if we reflected carefully, we should be able to anticipate emergencies that will require exceptions.)

(15.) In favor of this interpretation is the fact that in the *Critique of Practical Reason* Kant suggests that the universal law formula can only be applied through a "typic," the idea of a natural order, that mediates between the abstract moral law and our concrete situation (1997, pp. 59–60 [5:69–70]). In addition, Kant's examples in the *Groundwork* all refer to laws of nature. Some evidence to the contrary is that Kant does not refer to laws of nature in the formula that (in his review of formulas, 2002, p. 237 [4:436–437]) he says is best to use in moral assessment. In addition, Kant sometimes gives examples that suggest that the relevant question is whether we can conceive and will our maxim as one that it is permissible for everyone to act on. (p. 204 [4:403]).

(16.) See Paton, 1958, chs. 14, 15. Under FUL, Paton says, we ask whether "we can will our maxim as an instance of a principle valid for all rational beings and not merely adopted arbitrarily for ourselves" (p. 146). With FULN, by contrast, "we put ourselves imaginatively in the position of the Creator and suppose that we are making a world of nature [i.e., a system of natural purposes] of which we are apart" (p. 146).

(17.) An example of this type of interpretation is developed by Pogge, 1998.

(18.) We must try to conceive everyone adopting the maxim and acting on it *in conditions where it is applicable*. The conditions should be given in the maxim, which has the form "In conditions C, I shall do X …" Presumably, when we conceive a maxim as a *universal law of nature*, we assume that some appropriate conditions will occur and so the maxim will be acted on. Otherwise, a maxim would pass the first test if, though no one could act on it, everyone could "adopt it" (perhaps in the form "In C, I shall X if I can"). To

understand the test that way would undermine Kant's most persuasive example, that of a lying promise to repay a loan (Kant, 2002, p. 223 [4:422]).

(19.) Brandt, 1959, pp. 27–35, among others, noted this. An example would be the fact that people would generally remember those who did not repay loans and would be reluctant to trust those persons again.

(20.) Paton, 1958, pp. 146–156.

(21.) See Nell, 1975. In this early work, she maintains that one who wills a maxim thereby, if rational, wills the normal foreseeable consequences of what she wills. For example, if I will for everyone to adopt and act on my maxim to rob a bank for personal enrichment, then, if rational, I thereby also will the normal foreseeable consequence that bank security will become prohibitively tight. To will this, of course, is inconsistent with my original intention to enrich myself by bank robbery. I express some doubts about whether this account is satisfactory in my review of *Acting on Principle* (Hill, 1979).

(22.) Their argument would assume that the universal law formulas provide a sufficient, not merely necessary, condition of moral rightness. Wood, 1999, pp. 76–110, for one, argues that they do not provide such a criterion and were not meant to.

(23.) The term "rationally" has apparently acquired a narrower sense in our times than Kant's sense, which arguably encompasses more broadly what we might prefer to call "reasonable." Thus, in my view, Kant's test should be construed as about what we can rationally *and reasonably* endorse as universal policies. Adding the term suggests, not inappropriately in my view, that the reasons relevant to whether we can will maxims as universal law are not restricted to logical inconsistency and formally incompatible intentions. However, it would be contrary to the structure of Kant's moral theory simply to rely on intuition unrelated to any version of the Categorical Imperative for the standards of what can be reasonably willed as universal law.

(24.) Although as human beings we do not always "will" what we would if perfectly rational, Kant identifies our "will" with certain practical capacities we have as rational beings. He repeatedly says that acting on maxims that we cannot will as universal law implies "contradiction" or "conflict" in the will that would be absent if we were more fully rational.

(25.) For example, to argue that it is unreasonable to will a certain policy as universal law because that would undermine certain traditional values, we must independently show (by Kantian principles) that we ought to preserve those traditional values. Most contemporary Kantians seem to accept that all other moral norms must be "derived," at least in a broad sense, from the Categorical Imperative and noncontroversial principles of instrumental rationality, but one could treat Kant's principles as a merely as framework of general moral considerations that constrain all other moral considerations but need to be supplemented by them. We would naturally wonder, of course, what is the source of these

other moral considerations, why we should accept them, and whether disagreements about them can be resolved.

(26.) Kant himself suggests that in ordinary cases we come to our moral problem already having a stock of relevant moral beliefs, and if so, what reflection on the universal law formulas may do is simply to highlight the discrepancy between these general beliefs (that we readily apply to others) and our own proposed action (Kant, 2002, pp. 224–225 [4:424]). Here, asking whether we can will our maxims as universal laws is a way of considering whether our proposed act fits ideas that we already have about what everyone may, and may not, do. Clearly, there is a "contradiction the will" of someone who, without good reason, makes herself an exception to general principles that she accepts as reasonable.

(27.) Singer, 1961, p. 235, for example, argues this.

(28.) Herman, as we shall see, has an argument against the miser that does not rely on the humanity formula, but I am not convinced that contemporary Kantians should follow her in this.

(29.) Here I follow roughly Herman's line of argument (1984, pp. 577–602).

(30.) Kant holds that it is a duty to adopt the maxim, not merely to aid those in distress, but more generally to promote the (permissible) ends of others. Regarding how much one must do to promote others' happiness, and when and to whom one must do it, however, the duty is not determinate. Beneficence is a "wide, imperfect" duty.

(31.) The latter implausibly stringent principle is a little-noticed consequence of Ross's famous system (1930, ch. 2) of "prima facie duties," at least as long you think of your own happiness as *yours* rather than thinking of it impartially as merely the good of some person. Cummiskey, 1996, esp. chs. 5, 6, argues that Kant is committed to a slightly weaker, but still incredibly stringent, principle of beneficence, namely, that, absent other duties, we must maximize the general good (including happiness and rational development), taking all persons (including yourself as one) into account.

(32.) Earlier I mentioned a few guidelines for constructing a maxim that may help to set aside certain counterexamples, but I doubt that, even if supplemented with other suggestions, such guidelines are sufficient.

(33.) To derive this more general duty from universal law formulas, it seems that we would need to show that unless we adopted a maxim of promoting the (permissible) ends of others, we would act on some maxims that we cannot will as universal law.

(34.) Kant apparently treated these expressions as equivalent, but it may matterwhich expression is taken as primary. Kant frequently contrasts our humanity (or rational nature) with our animality, as different aspects of our nature as human beings. In his *Religion within the Boundaries of Mere Reason*, 1991b, he contrasts our animality with both our personality (or moral predisposition) and our humanity (other features of rational agency). Earlier I argued that we should understand the expression "treat rational nature in persons as an end in itself" as simply as saying more specifically what is meant by "treat persons as ends in themselves." It would be a mistake, however, to equate these with a command to do whatever promotes the greatest possible development and use of reason, for Kant is clear that an end-in-itself is not an end in the sense of a goal or something to be produced. See my 1980.

(35.) Not all commentators agree about whether the autonomy that grounds human dignity is our capacity and disposition to morality (i.e., giving ourselves moral laws as standards) or the actual realization of this capacity (i.e., willing always to conform to the moral law). Most commentators favor the former interpretation, but some passages suggest the latter. See Dean, 1996, pp. 266–288.

(36.) Since dignity is "above price," it obviously cannot be legitimately exchanged or sacrificed for commodities or pleasures; but, apparently though less obviously, Kant means that dignity has no "equivalent" even among other things with dignity. Thus, insofar as persons have dignity, they have an incalculable value that prohibits justifying one violation of human dignity by the thought that it would prevent two or more similar violations. This does not necessarily imply, however, that when we cannot save everyone's life in an emergency it would be wrong to do what would save more rather than fewer lives. I discuss this problem in my 1992c.

(37.) There is an often-noted tension between these passages, because the first tells us that ends-in-themselves are to be conceived only negatively, while the second prescribes a "positive agreement" with the idea.

(38.) See Singer, 1961.

(39.) See Singer, 1961, p. 235. See also Wolff, 1973, p. 176.

(40.) This is the position of Donagan, 1977, pp. 57–74. Donagan, however, does not claim that the first principle is self-evident or indubitable. In a rational ethical theory principles can be presented in hierarchical form, representing more specific principles as deductive conclusions from more abstract and comprehensive higher level principles; but, in considering reasons to accept the system, we may have initially more confidence in the more specific intermediate principles. Thus we may treat the fact that the "first principle" leads to just these principles, and not to conclusions we find unacceptable, as some confirmation that the first principle, as stated, is correct. See Donagan, 1993. Wood gives the humanity formula a crucial "grounding" role in his interpretation of Kant's basic

ethical theory, but does not give it as exclusive a role as Donagan does. See Wood, 1999, pp. 111–155.

(41.) There are as variety of possible views of this sort. For example, one may think that the humanity formula only makes explicit what was presupposed in the formula of universal law, thus helping to clarify or interpret it. One may think that they are independent principles that are at best practically equivalent, in that they yield the same conclusions about cases. Or one may think that these two formulas each express certain aspects of a more comprehensive principle that is expressed in a later formula (or must be constructed from all of the formulas). The latter is more or less the position I propose as a reconstruction, but not strict interpretation, of Kant's views.

(42.) See O'Neill, 1989, pp. 137–140.

(43.) This was the theme of my early essay "Humanity as an End in Itself," 1980. Cummiskey, 1996, also begins from an interpretation of the humanity formula based on the idea that to be an end-in-itself is to have dignity, as opposed to price.

(44.) Donagan's account, 1977, treats the formula as a principle of respect, but in *The Metaphysics of Morals*, 199a, Kant's discussion of respect for others is concerned only with certain aspects of morality (contrasted with beneficence, gratitude, etc.), not as a comprehensive moral guide. Similarly, Kant seems to treat self-respect as a more specific requirement than the Categorical Imperative, though one could argue that all duties to oneself are "really" requirements of self-respect.

(45.) Although for purposes of developing Kantian ethics I suggest that an alternative interpretation in terms of dignity is more promising, I think that the account proposed here has some merit, at least as an interpretation of Kant's remark that ends must be valued *as able to share our end*. In "Donagan's Kant," 1993, I suggest this leads to a "thin" reading of the humanity formula, in contrast with an earlier "thick" reading that emphasized the incomparable value of rational agency. See my *Respect, Pluralism, Justice: Kantian Perspectives*, 2000, pp. 148–151.

(46.) For example, my "Humanity as an End in Itself," 1980.

(47.) See Cummiskey, 1996.

(48.) See Wood, 1999, p. 141. I turn to Wood's view shortly. See also Korsgaard, 1996a, pp. 106–132, esp. 275. Both Korsgaard and Wood emphasize our value-conferring status as beings who rationally set ends for ourselves.

(49.) For example, Donagan's principle regarding lying is this: "Even for a good end, it is impermissible for anybody, in conditions of free communications between responsible persons, to express an opinion that he does not hold" (1977, p. 89).

(50.) See Williams, 1973, pp. 77–155.

(51.) For example, the humanity formula can ground positive duties, express the basic value that "grounds" our moral judgments, and is not liable to generate the bizarre, counterintuitive results that the universal law formulas may yield when maxims are described in certain ways.

(52.) Kant says that we have a duty to promote the *happiness* of others, and he typically understands happiness as a subjective idea of lasting contentment or realizing all one's desire-based ends (not the same idea as the classic idea of human flourishing). I discuss the contrast in my 1999.

(53.) Donagan, 1977, p. 85, includes persons orphaned, grieving, injured, chronically ill, blind, deaf, senile, etc.

(54.) See my 1993, pp. 46–49.

(55.) This idea is proposed and illustrated in my 1992c.

(56.) This suggestion is developed further in the next section and in my *Dignity and Practical Reason in Kant's Moral Theory*, 1992a, chs. 2, 10, 11, and *Respect, Pluralism, and Justice*, 2000, chs. 2, 4, 8.

(57.) In the kingdom of ends, the lawmakers are conceived as having a system of personal ends, even though in thinking of the kingdom we "abstract" from their content. Having ends that one sets oneself is a necessary feature of being a rational agent, though these ends can vary among different kinds of rational beings (e.g., human and nonhuman) and among individuals of the same kind. This has implications about themotivations of the rational legislators, even when (as if behind a "veil") they abstract from the content of their ends. For example, they would tend to favor whatever promotes everyone's ends and frustrates no one's, and whatever establishes conditions under which all can effectively pursue their ends. Because, as members of the kingdom of ends, they regard the humanity of each member as an end-in-itself, then they have a general motivational stance that will encourage promoting others' ends but rule out certain kinds of ends as unworthy.

(58.) Kant's belief, or perhaps faith, that virtually all competent adult human beings implicitly acknowledge his noninstrumental moral principles may be hard to sustain in our times, but it can be seen at least as a morally reasonable working presumption in the absence of compelling evidence regarding particular cases. In addition, we may see Kant as offering a model of rational agency (with related conceptions of "reasons," "reasonable," etc.) that (arguably) is presupposed in common thought, and this can be valuable, even if not every member of our species turns out to satisfy its criteria for being a "rational agent." Kant's claim that *rational* agents at least implicitly acknowledge his basic moral principles, I take it, is a substantive normative thesis, not an empirical hypothesis and not a proposition "analytically" true by virtue of the ordinary meaning of the word "rational" (or a special meaning that Kant stipulates). Contrary to some influential Kant scholars, I think that Kant's primary argument for his thesis is that

Subscriber: King's College London; date: 24 November 2018

common moral consciousness (especially the idea of "duty") *presupposes* that we are practically rational agents who, as such, acknowledge his basic principles. This, of course, is not an argument that would move anyone who lacked the sort of moral consciousness that Kant took for granted.

(59.) The kingdom of ends is said to have a "head" or "sovereign" who apparently legislates the same laws for the same reasons as the members. The head is not "bound" by the laws, presumably because, lacking in all needs and inclinations, the head is never tempted to deviate from rational principles. As Kant earlier says of a "holy will," the language of imperatives, necessitation, and "ought" is inappropriate when there is no possibility of misconduct.

(60.) The "sovereign" or head (*Oberhaupt*) of the kingdom is technically like a state sovereign as traditionally conceived, in that the sovereign makes laws but is not *bound* by them, *subject* to them, or under *obligation* to follow them. This does not mean, however, that the sovereign may, or even can, act contrary to the laws that all members, including the sovereign, legislate. The point is that the sovereign, a "holy will," necessarily does whatever is rational by the sovereign's own nature, and so talk of the sovereign's obligations is out of place. Importantly, the members would legislate the same laws (because they are rational) and have the same obligations even if there were no sovereign.

(61.) The text is ambiguous at several relevant points. For example, it is not clear whether Kant means that "each of [the three main formulas] by itself uniting the other two within it" (Kant, 2002, pp. 236–237 [4:436]) or "one of them by itself containing a combination of the other two" (Kant, 1964, p. 103 [4:436]). If the latter were right, the one that combines the others would clearly be the kingdom-of-ends formula. Again, after Kant says that "[a]ll maxims have … a form, … a matter, … and a complete determination," it is not entirely clear whether the subsequent expressions of the moral law are meant as characterizations of the form, matter, and complete characterization of themoral law (the *maxim* of a good will) or requirements regarding the form, matter, and complete determination of ordinary maxims of action (Kant, 2002, p. 237 [4:436]).

(62.) A reasonable extension of Kant's idea, I think, would have to view its application as proceeding in several stages, in a way to be explained later. At the first stage, we would imagine all rational agents legislating for all rational agents and so not tailoring their principles for any local circumstances.

(63.) Wolf, 1973, p. 183, for one, endorses this interpretation.

(64.) The primary similarity to note between Kant and Rawls here is in the "constructivist" structure of the theories. There are other significant similarities, of course, but there are also very substantial and important differences. I have discussed the proposed legislative model, its relation to Rawls's work, and the problems it raises in several essays, and only a few basic points will be sketched here. See my *Dignity and*

Practical Reason in Kant's Moral Theory, 1992a, chs. 3, 4, 11, and *Respect, Pluralism, and Justice*, 2000, chs. 2, 4, 8.

(65.) This is not to say that we take our moral standards from our historical circumstances. The point is that in using our basic standards to reach particular conclusions, we need to take into account the facts about our historical condition, at least insofar as they are relevant under those basic standards. We should not assume at the outset that we know the scope of the principles we can justify from basic values, how much they are open to exception, and even whether they are general enough to be useful guides. This is part of what would need to be worked out.

(66.) I discuss this question in more detail in "Meeting Needs and Doing Favors," in my 2002, ch. 7.

(67.) This, in any case, is how Kant views wide imperfect duties, such as beneficence, in *The Metaphysics of Morals*, 1991a.

(68.) I discuss these and some other objections in *Respect, Pluralism, and Justice*, (2000), esp. ch. 2, and *Human Welfare and Moral Worth*, 2002, ch. 3.

Thomas E. Hill Jr.

Thomas E. Hill Jr. is Kenan Professor of Philosophy at the University of North Carolina, Chapel Hill. Kenan Professor.