Kantianism for Consequentialists

SHELLY KAGAN

Kant's moral philosophy represents one of the most significant approaches to the foundations of ethics. For obvious reasons—including the simple fact that Kant offered no distinctive name for his general approach to ethics—views of this same, basic sort are typically known as Kantian. But this common practice, natural as it is, carries with it an obvious danger as well: there is a temptation to assume that Kant himself is the last word on Kantianism, rather than merely being an important advocate of this *sort* of view. This can lull us into overlooking the possibility that in various places Kant may have been mistaken about the implications of Kantianism; and it can also make us feel needless pressure to reconstruct Kantianism in precisely the terms in which Kant himself presented it. As a result, we may narrowly focus on the details of Kant's particular views, at the expense of appreciating the fuller significance and general interest of Kantianism. (In contrast, we are quite used to thinking of Bentham, Mill, and Sidgwick as merely being leading representatives of the general *utilitarian* approach, without thinking that any one of them has the last word on utilitarianism itself.)

In this paper I want to discuss one significant strand of Kantianism in ethics. I focus, in particular, on certain ideas put forward in the *Groundwork for the Metaphysics of Morals.* But I must emphasize the point that the ideas I will be discussing are primarily put forward here as being Kantian, rather than Kant's. The position I will be discussing is certainly inspired (at a minimum) by Kant's own discussion in the *Groundwork*, and I will periodically turn to the text of the *Groundwork* itself for guidance and comparison. But this essay is not intended as a piece of Kant scholarship. Rather, it is intended as a contribution to understanding Kantianism. Indeed, because of this, I will hereafter refer to *k*antianism (rather than *K*antianism) where the lower case "k" is intended to mark the idea that I am primarily interested in the *type* of approach that Kant represents, rather than Kant exegesis per se. What I want to do, then, is to sketch the basic elements of a possible kantian approach, and indicate why I think the view has abiding significance for moral philosophy.

My primary goal is expository. I hope to say enough to make it clear why kantianism is worth taking seriously—even by those who may, at the end of the day, choose not to accept it. It is not my intention to offer anything like a *full* presentation of kantianism (we will only be considering a few of the main ideas discussed in the *Groundwork*), nor is it my intention to offer anything like a full *defense* of kantianism. While I hope to say enough to show why one might find kantianism attractive and plausible, the arguments I offer are only rough sketches, and many important objections will go unanswered (or unmentioned).

I have a secondary goal as well, reflected in my choice of title. Kant himself believed that kantian foundations supported a deontological rather than a consequentialist normative theory.[1] Since most philosophers have assumed that he was right about this, those sympathetic to consequentialism have typically had little interest in understanding kantianism. But in fact it is far from clear whether Kant *was* right about this.[2] So I hope to offer an account of kantianism that consequentialists may find congenial. In any event, if I am right in thinking that kantian foundations are themselves fairly plausible, then it behooves those who want to reject those foundations to identify exactly where they think those foundations go wrong. (Of course, given the obscurity of much of Kant's writing, it may not be surprising that few consequentialists have actually attempted to do this.) Accordingly, I want to offer a guide to kantianism that may be of particular use to consequentialists.

But this further goal is indeed only secondary. My primary purpose is to sketch the main lines of a potentially attractive version of kantianism. Questions about the particular normative implications of kantian foundations can be put aside until we have a better handle on the kantian foundations themselves.

I. Autonomy and the Formula of Universal Law

Where then should we begin? Kantianism begins with freedom. More particularly, it begins with the fact that we are free, and with an account of that freedom. So we must begin with that account. (It is worth noting, however, that the *Groundwork* itself does not begin with the idea of freedom, but rather works backward toward it, arguing in the first two sections that if there is to be such a thing as morality, then we must be free—that freedom

is the basis of morality. Unfortunately, in the *Groundwork* itself Kant says rather little explicitly about how exactly we are supposed to be able to move from the assumption of freedom back "up" to morality (see G 4:446–47). Thus we must depart from Kant exegesis almost immediately.)

Kantianism begins with freedom. But I think we will better understand the relevant notion of freedom if we begin instead with rationality. What, exactly, is it to be rational?

Suppose we start with theoretical rationality. As a theoretically rational being, I am capable of examining my various beliefs and seeing whether it makes sense for me to hold them. Thus, in the first place, I have *standards* for evaluating beliefs, in the light of which I can ask whether or not I am justified in holding a given belief. I might, for example, appeal to various principles of logic, discovering that some of my beliefs commit me to accepting still other beliefs; or I might appeal to various rules of scientific methodology, finding that, given the available evidence, I am unjustified in accepting some further belief. But rationality in the theoretical domain goes beyond the mere *evaluation* of my beliefs: I can *change* my beliefs in light of my judgments concerning the extent to which they meet (or fail to meet) the relevant standards. Normally, that is, when I see that the evidence better supports one claim rather than another, my beliefs change accordingly. Roughly, then, theoretical rationality consists in my ability to evaluate my beliefs in light of the standards relevant for evaluating beliefs, and to alter my beliefs in the light of those evaluations.

Practical rationality is similar. As a practically rational being I am capable of examining my various desires, goals, intentions, actions, and the like, so as to see which of these make sense in the circumstances. Here too, then, I have standards in terms of which my plans can be evaluated, goals assessed, actions endorsed or criticized. Nor are these various practical elements merely subject to evaluation; I can *change* my goals, my intentions, and the like, in light of my judgments concerning the extent to which these meet (or fail to meet) the relevant standards. Thus practical rationality consists in my ability to evaluate actions, intentions, and so forth, in light of the standards relevant for these, and to alter these elements in light of those evaluations.

Generalizing, then, we can say that rationality—whether practical or theoretical—consists in the ability to evaluate beliefs and acts (and so forth) with an eye to whether they meet the relevant standards, and to alter our beliefs and acts in light of those evaluations.

In this way rationality goes beyond mere intelligence. Nonhuman animals, I presume, also have beliefs and desires, and act in a way that is often appropriate to their circumstances. Some animals may well be extremely adept at achieving goals and forming appropriate beliefs about their environment. Thus they display varying (and perhaps considerable) degrees of intelligence. We could say that *intelligence* consists in the ability to produce beliefs and actions that in point of fact are appropriate (that is, conform to the relevant standards); animals are often intelligent in this sense. But only rational creatures are capable of *articulating* the standards against which beliefs and actions are to be evaluated, and only rational creatures are capable of consciously *comparing* beliefs and actions (real or imagined) against those same standards (G 4:412, 427).

It is worth emphasizing as well the point that as rational beings we are capable of *rejecting* the beliefs and actions (and the like) that don't meet what we take to be the relevant standards. We modify our behavior and our beliefs in light of what we think appropriate. For example, we are not normally *forced* to act on desires that we happen to have, when we conclude that such desires don't make sense, or that acting on them in present circumstances would be inappropriate (by whatever standards we here take to be relevant). In this way, too, rational beings are different from merely intelligent animals. For it seems plausible to view animals as mere "playthings" or "puppets" of their desires—incapable of evaluating them, and thus incapable of rejecting them. In contrast, rational beings are in an important sense *free*: if we conclude that a given desire makes no sense (perhaps we recognize that it was based on what we now see to be a mistaken belief) or that a given intention is inappropriate, we are free to step back from that desire or intention, and to refuse to act on it.

Of course, the simple fact of the matter is that humans are not *perfectly* rational. At times we misapply our own standards and fail to see that a belief cannot be justified (given the relevant standards). Or we may find ourselves incapable of *abandoning* certain beliefs, even though we can see that these beliefs are not in fact justified. Similarly, at times we may find ourselves giving in to desires, even though we see full well that acting on this desire, in this situation, doesn't actually make sense, or is otherwise inappropriate. Thus we are, at best, only imperfectly rational. Still, it would be implausible to suggest that we are not rational at all (in this sense), for we clearly are capable of articulating standards for evaluating beliefs and actions, and we are typically capable of evaluating our beliefs and actions in the light of those standards; and often, at least, we are capable of modifying our behavior and beliefs in the light of those evaluations. Humans may not be perfectly rational, but we are rational nonetheless, even if only imperfectly so.

The account of rationality that I have been sketching is, indeed, only a sketch. But even so, it remains significantly incomplete, in that I have not yet drawn attention to an important further fact: not only are we capable of articulating relevant standards, and evaluating and modifying beliefs and actions in light of those standards; the standards *themselves* are things that we can evaluate and modify. That is, for any given standard that I might use to evaluate a belief or an act (or an intention, and so forth), I can ask of the standard itself whether *it* makes sense, whether *it* is indeed an appropriate standard to be used in this way in these circumstances. In effect, I can ask whether the given standard itself meets the standards (whatever they are) relevant for evaluating *standards*. And armed with these evaluations, I can in turn reject any given standard, modify it, or replace it. Thus, as a rational being I am free not only to reject, modify, or endorse my various beliefs and actions—I am also free to reject, modify, or endorse the standards I appeal to in evaluating beliefs and actions. I am not forced to accept and appeal to standards that do not make sense to me or that seem unjustified or inappropriate. I am free to alter the standards as I see fit.

And the same is true, of course, with regard to the "second order" standards that I may use to evaluate the "first order" standards. These higher order standards can themselves be subject to critical evaluation: I can ask whether the standards I use for evaluating standards are themselves appropriate, whether they themselves meet the relevant ("third order") standards (whatever I may take these to be) for evaluating such (second order) standards. And I can modify these higher order standards as seems appropriate in light of these further evaluations. And so on, and so forth, all the way up (or all the way down): no standard is itself forced upon me, no standard is immune to potential criticism or evaluation. I am free, in principle, to evaluate any standard whatsoever, to ask whether it makes sense to me, whether it is indeed an appropriate standard to use. The principles or standards by which I evaluate beliefs and actions are themselves subject to rational assessment and open to modification or rejection. Put another way, the rules of rationality are not forced upon me (against my will, as it were): I need only appeal to standards that make sense to me, that seem appropriate in light of whatever principles, rules, or standards I endorse.

Our examination of the nature of rationality has thus led us to an important insight. The rules or standards to which I appeal in rationally assessing beliefs and actions are themselves subject to rational assessment, and at no point need I simply accept a relevant rule or standard as simply given—from "out there," as it were, forced upon me despite its making no sense. On the contrary, the relevant rules or standards need only be accepted if they, too, make sense in light of whatever rules and standards I reasonably accept. We could put the point this way: the laws of rationality are not forced upon reason from the outside. Rather, reason is free to reject those standards (at whatever level) that do not make sense to itself. Reason is its own last court of appeal. It chooses what standards to obey. In short: reason is *autonomous* (G 4:440).

The fact that reason is autonomous in this way is certainly not altogether obvious. Indeed, Kant believed that previous moral philosophers had failed to recognize the autonomy of reason, and certainly had failed to appreciate the implications of reason's autonomy for ethics (G 4:432–33). Most moral philosophies have been founded in heteronomous conceptions of reason, where some ultimate principle of reasoning is simply taken as "given" (from outside reason's control) and beyond question (G 4:441–44). But kantians believe that since we are autonomous (insofar as we are rational), all such approaches to ethics must fail. If there is to be any hope for a sound foundation for ethics, it must take account of our autonomy.

In the account I have been sketching, the ideas of reason, freedom, and autonomy are tightly connected. Clearly, much more needs to be said, both in defense of the general kantian picture I have been presenting, and by way of further clarification of the three related concepts. But I am going to restrict myself here to two quick remarks.

First, our analysis of rationality has led us to a picture of rational beings as free. So eventually the kantian must confront the question of whether the freedom that we take ourselves to have (as rational beings) is genuine or a mere illusion. Kant himself postpones the discussion of this issue until the third section of the *Groundwork*, and even there the discussion is cursory. In this essay I shall make no attempt whatsoever to pursue this question.[3] I believe it plausible to hold that we are free, in the relevant sense, but I won't attempt to defend this claim here. And so, along with Kant in the first two sections of the *Groundwork*, we can view the rest of our discussion as taking the form of a conditional: if we *are* free, what follows?

Second, I want to say a word more about the concept of autonomy. Kant typically expresses the thought that reason is autonomous by saying that reason is the *author* or *source* of the rules and standards used by reason (e.g., G 4:431). But it is not clear that our concept of rationality can take us quite this far. Suppose we grant the kantian that the freedom involved in rationality means that there are no sound or valid standards for rational assessment that cannot themselves withstand the scrutiny of rational assessment. This would mean that there are no valid rules of reasoning that reason doesn't itself "accept," or "will," or "approve." We might capture this idea by saying that reason must itself "sign off" on any purported rules of rationality that are themselves to be binding upon reason. (There are no rules binding upon reason that reason wishes itself free of, no rules that it considers unreasonable rules.) But is it also true that we must think of reason as the *author* of these rules (the *ground* of their validity)? Kant apparently thinks so, though it is not clear why. Perhaps (and this is sheer speculation) he believes that it is inexplicable how reason *could* have this kind of veto power over rules of reasoning (so that no rule it disapproves of is valid) unless reason is itself the *source* of the validity of the (valid) rules of reasoning. This claim is not completely unattractive, and so I shall follow Kant here in speaking of reason as the source or author of its own rules. I believe, however, that this further claim is not strictly needed by the kantian. So long as it is conceded that reason's autonomy means that reason must "sign off" on any principles of rationality if they are indeed to be sound — that no standard for rational assessment is valid unless reason itself can approve of it — the kantian has, I believe, all that he needs.

Now kantians believe that *given* the autonomy of reason, certain implications fall out concerning the rules or standards that reason can give to itself. In particular, they believe that once we recognize the autonomy of reason, we are committed to accepting a certain fundamental rule — the *universal law* formulation of the categorical imperative (FUL). Here is a possible reconstruction of the main line of thought.

Whenever I act, my acting presupposes that there is reason to do whatever it is that I am doing, that my act makes sense in the given circumstances. In effect, each action presupposes some rule or principle (though not necessarily the same rule from act to act) that endorses the act, a rule in the light of which the act can be seen as reasonable. Typically, of course, these underlying principles or rules will only be implicit, but were we to make them explicit, they might say something along the following lines: under such and such circumstances, given such and such desires or such and such goals, there is reason to act in such and such a way. As I say, we rarely make such rules explicit (and even less frequently attempt to state them fully and with care), but whenever I act, I presuppose some such rule — a rule which, if sound, would validate my action, by showing why it is that I have reason to do whatever it is that I am doing. (In many cases, of course, one acts spontaneously, or simply "goes with the flow." But in such cases, presumably, the principle implicit in one's act is precisely one that endorses acting spontaneously in circumstances of this sort.)

So when I act, I presuppose a rule or principle that claims that I have reason to do what I do (given the circumstances, and so forth). But which rules should I act on? This much seems clear: I should only act on rules that are themselves *valid.* (The precise term of commendation used here isn't important for our purposes. We could equally well talk of those rules that are sound, or legitimate, or good, or reasonable.) I should only do what it truly makes sense for me to do; so I should only act on those rules that are themselves correct in their claims about what it is that I have reason to do. I should only act on those principles that are valid.

But given that I am autonomous, the rules are up to me. Valid rules are valid by virtue of my signing off on them, by virtue of my approving of them as a rational being.

So this means: I should only act on rules that I can sign off on. I should only act on rules that I can rationally choose to be rules. Put in slightly different terms: I should act only upon rules that I can (rationally) *will* to be rules.

But rules are *laws.* They tell everyone what to do (or believe, or intend, and so on) in relevant circumstances. They say, for example, that in such and such circumstances, given such and such desires, one has reason to perform an act of such and such a type. But this means (if the given rule is valid) that *everyone* has such a reason—provided that they have the relevant desires and find themselves in the relevant circumstances. Of course not everyone will necessarily find themselves in the relevant circumstances, or with the relevant desires—but it is true of everyone that *if* they were in the relevant circumstances (and so forth) then they *would* have reason to perform an act of the relevant sort. Rules are *universal*, providing the same reasons (under the relevant circumstances) to everyone.

So we can restate our earlier conclusion. Instead of saying that I should act only upon rules that I can (rationally) will to be rules, we can say: I should act only on those rules that I can (rationally) will to be universal laws.

This is Kant's formula of universal law, though his own favored statement of it makes use of a piece of jargon. Kant typically talks about *maxims*, which for our purposes we can take to be first person statements of intentions ("I will perform such and such an act in such and such circumstances, given such and such goals"). Each such maxim corresponds to an implicit principle ("if one is in such and such circumstances, with such and such goals, then one has reason to perform such and such an act"), and so we could restate the formula at which we have arrived as follows: I should act only on those maxims, where I can (rationally) will that the corresponding principle be universal law. Simplifying a bit further still, we can say: act only on those maxims that I can will to be universal laws. And this is

exactly what Kant tells us. Here is his own statement of the formula of universal law:

> FUL: "Act only in accordance with that maxim through which you at the same time can will that it become a universal law" (G 4:421).

Kant's decision to state FUL in terms of maxims rather than the corresponding principles carries certain risks, for one can normally state one's intentions in a way that only gives a partial indication of what one takes oneself to be doing, and why it seems to make sense. Thus, for example, if my intention is to close the door to keep out the person attacking me, so as to save my life, it will normally be correct to say, as well, that I intend to close the door. But if we then focus on "I will close the door" as a statement of my maxim, we will have no idea (or at best a poor idea) of why I think it makes sense to do this in the present circumstances, and thus no idea (or at best a poor idea) of just what the corresponding principle is supposed to be that I am to examine so as to see whether I can indeed will it to be universal law. These problems could have been avoided had Kant stated FUL directly in terms of examining complete statements of the underlying principles. But so long as we bear in mind that the real question is always whether a purported reason-giving principle is indeed one that we can rationally will to be universal law, we should be able to make use of Kant's own formulation without too much confusion.

Now the argument I have just sketched moves from our autonomy to FUL, a requirement to act only on certain types of maxims (in Kant's formulation). But if this argument is sound, then the resulting requirement should apply equally to *everyone*, that is, to every rational being. For if reason is autonomous, and autonomy yields FUL, then FUL is binding upon all rational beings. That is to say: all rational beings should obey FUL; they *must* do it if they are to act rationally. We can express this point in kantian jargon by saying that FUL is a *categorical imperative* (one binding upon all rational beings; see, e.g., G 4:432). Of course this does not mean that all rational beings *will* obey FUL. As we have already noted, humans, at least, are only imperfectly rational, and thus may often fail to conform to FUL, sometimes knowingly. But everyone *should* obey FUL: they have reason to do so, based on the mere fact that they are rational. If the argument is sound, then FUL is a categorical imperative.

Kant says there is exactly one categorical imperative, though it has several equivalent formulations (G 4:420–21, 436). FUL is supposedly only one of the different ways of stating this single imperative. Another of the formulations, the formula of autonomy (FA), goes like this:

> FA: "the idea of the will of every rational being as a will giving universal law" (G 4:431).

Note that Kant doesn't even bother to state this version in the form of an imperative at all! Presumably, however, what he is most concerned to impress upon us here is the idea that it is autonomy (the fact that reason is the source of its own laws) that provides the basis for FUL: given the former, we are led to the latter. The argument I have been sketching tries to make good on this thought. (To get full equivalence, of course, we would also need to go on to argue as well that given FUL we can derive an imperative along the lines of "Act autonomously!" or "Act in keeping with your autonomy!" I won't attempt to argue that here.)

But is the argument sound? Can we actually derive FUL from the mere assumption of reason's autonomy? I am not sure. Doubtless several steps of the argument could be questioned, but the most important issue, I believe, is this. Is it really true that the only rules or standards that I could autonomously will are *universal*? Must the reason-giving principles I endorse be principles that would equally give *everyone* a reason? Putting the same point in a slightly different way, is it really true that the only rules that I could freely give to myself are rules that make similar prescriptions for everyone? Unless something like this is true, then all that autonomy will demand is that I act on maxims that I can (autonomously) will. We won't have a requirement that I act only on maxims that I can will to be universal law. And so we won't have made it all the way to FUL. So we need to ask: is it really true that the only principles I can autonomously give myself are universal?

Now it might seem that the answer to this question is obvious. For it seems obvious that I can (and should!) endorse principles that recognize that what *I* have reason to do normally differs from what *you* have reason to do. For example, I may have reason to eat right now, while you do not.

In thinking about this question, however, it is important to bear in mind the point, already noted, that the requirement that the reason-giving principles be (ones that I can will to be) universal laws only amounts to a requirement that people *in the same circumstances* have the same reasons. Thus, universality here only amounts to the requirement that *if* someone else were in the same circumstances (that is, whatever the principle takes to be the relevant circumstances) then they too would have reason to perform the same kind of act. (And it should be noted that, depending on the given principle, the relevant circumstances may well include a specification of the person's desires or goals as well as more "external" circumstances.) So even if the principles I give myself are universal, this doesn't mean that everyone has reason to do the same specific types of acts, for people will still find themselves in differing circumstances.

In typical cases, at least, when we find ourselves thinking that one person has reason to do something that another person does not, this will be because we think there is some relevant difference in their circumstances—and a full specification of the relevant reason-giving principles will take note of these circumstances. Thus, for example, I may believe that I should eat, while you should not, but this may be because I believe that only hungry people should eat, and I recognize that you are not hungry. (Or perhaps I believe that people on diets shouldn't eat between meals, and you are on a diet and it is between meals; or that you need to get to a class, and I do not, and so on.) Despite initial appearances to the contrary, then, the underlying principle will actually be universal: anyone who is similarly situated (with regard to hunger, dietary needs, availability of food, more pressing demands, and so forth) will have similar reason to eat (or not). If this is right, then at the very least most of the principles I can actually sign off on will indeed be universal laws in the relevant sense.

But is it truly *impossible* for me to autonomously will principles that are not in this way universal? Can't I simply endorse a rule that says that *I* (but not others) should do such and such an act in *this* case (but not in other cases that are otherwise similar)?

Here I can only reply that when I honestly contemplate such irreducibly person specific or irreducibly case specific principles I find them virtually unintelligible. I cannot fathom the idea that I might have reason to do something in a certain kind of case, while you do not—even though there is not a single relevant difference between us. This is not to say that I can't imagine someone "stating" such a principle, nor do I mean to claim that I wouldn't understand what someone affirming such a principle would be attempting to do. Rather, I simply find that I cannot take seriously the possibility that such a principle would be one that merits endorsement. If in the circumstances someone has *reason* to act in a given way, then it seems to me that anyone at all who genuinely found themselves in relevantly *identical* circumstances would have reason to act in the same way. Which is to say, when I ask myself what sorts of reason-giving principles I can truly imagine autonomously giving to myself—fully accepting upon complete rational reflection—the only such principles are ones that are universal.

In my own case, then, if I am indeed to restrict myself to maxims that I can autonomously will, then I must restrict myself to maxims that I can will to be universal laws. Perhaps others differ from me in this regard. The idea seems just barely possible, though, again, I find that I can't take the thought seriously. As far as I can see, *any* rational being would find that the only maxims he could autonomously will would be maxims that he could will to be universal laws. And if this is indeed correct, then given the autonomy of reason something like FUL may well follow for all rational beings whatsoever. In short, FUL may indeed be a categorical imperative.

To be sure, other questions about this step of the argument could be pressed, and other stages of the argument could be challenged as well. So I would not want to claim that the validity of the derivation of FUL (from the assumption of autonomy) has now been established. But I hope that I have said enough at this point to make it clear why the kantian's appeal to FUL is a position worth taking seriously. The claim that reason is autonomous is, I think, a plausible one, and the further claim that autonomy yields FUL is not, I believe, one that can be easily dismissed. If nothing more, these claims are sufficiently plausible (even if one ultimately rejects one or the other of the pair) that what I have said should make it clear why many people have found FUL so compelling.

II. Understanding the Formula of Universal Law

Suppose, then, that we grant the kantians the validity of FUL (if only for the sake of argument). Even if we do this, it is hardly obvious how FUL is to be applied, how it is to be put to work. Nor is it the least bit obvious whether—as kantians believe—FUL has sufficient "bite" that it can be used to generate concrete moral guidance. So let us put aside further questions about the derivation of FUL, and turn instead to the question of what follows from it. Granted that I must only act on maxims that I can will to be universal law, how exactly am I to decide what to do?

The first thing to notice is that FUL itself doesn't actually provide us with maxims; it only serves to rule some of them out. We bring candidate maxims *to* FUL, to see whether they are acceptable. The point here is easy enough to grasp if we recall that maxims are, in effect, statements of what one intends to *do* in a given situation. What we should imagine then is that faced with the given situation, I have come up with some tentative plan of action, something that I propose to do (perhaps to serve some desire or goal I have). Armed with this tentative plan, then, I turn to FUL to see if it is legitimate to act on it. FUL is, in effect, a *test* of maxims: it tells me to act only on maxims that have a certain feature.

For the moment, let's leave the details of that test aside, and focus on the negative form of the imperative. FUL tells me to act only on maxims that pass a certain test. Thus, if some maxim *fails* the test, FUL commands me not to act on it. Notice, however, that although FUL tells me to act *only* on maxims that pass the test, it does not require me to act on all the maxims that *do* pass the test. Apparently, then, if a maxim passes the relevant test you *may* act on it (FUL, at least, won't rule this out); but absent any further argument, it seems, there won't be any *requirement* to act on the maxim. We must restrict ourselves to acting on maxims that pass the test, but among the maxims that do pass, which we choose to act upon is up to us.[4]

Suppose, then, that some maxim fails the FUL test (whatever, exactly, it turns out to be). What can we conclude? If FUL is indeed a categorical imperative, binding upon all rational beings, then we must conclude that it is forbidden to act on that maxim. But what follows in the alternative case, where the given maxim passes the test? Here we have to be more cautious. Obviously enough, if a given maxim passes FUL, then as far as FUL itself is concerned there is nothing objectionable about acting on the maxim. But we cannot yet safely conclude that it is indeed *permissible* to act on the maxim in question, because, for all that we have said so far, there might be some other imperative—beyond FUL—that must be taken into account as well. After all, even if kantians are right in thinking that reason's autonomy supports FUL, it doesn't yet follow that this is the only fundamental principle supported by our autonomy. Perhaps there are *additional* tests that must be passed as well. If so, then passing FUL will be necessary for permissibility but not sufficient.

Presumably Kant means to put this possibility aside with his insistence that FUL is the *only* categorical imperative. (Because of this belief, he typically refers to it simply as "the" categorical imperative, though as we have noted Kant also believes that this imperative can be stated in several different, though equivalent, ways.) But even if Kant *could* prove that FUL (in its various formulations) is indeed the only categorical imperative,[5] that wouldn't necessarily put the worry to rest. For what if there were additional, basic principles (that is, principles not derived from FUL) that, although not categorical, nonetheless validly applied in particular cases? Even if FUL is the only *categorical* imperative, nothing yet rules out the possibility that a maxim might pass FUL but nonetheless fail to pass these further (noncategorical) principles.

What the kantian needs to claim then (*regardless* of whether FUL is the only categorical imperative) is that even if there are any further valid principles (not themselves derived from FUL), it is not actually possible for a maxim to pass FUL but to violate these further principles. Happily, this may not be an implausible claim for the kantian to make. Imagine that a given maxim violates some such principle, P. Now given the autonomy of reason, any valid principle of reasoning, including P, must be one that I rationally favor. But if I truly continue to endorse P (even in light of its ruling out the maxim in question) then I cannot rationally favor any principle incompatible with P — including, in particular, the underlying principle corresponding to the maxim. Thus, given my acceptance of P, I cannot in fact rationally will the maxim to be universal law. That is, if the maxim violates P, it fails FUL as well.

What this means, then, is that even if there *are* additional principles (not themselves derived from FUL), so long as a given maxim does pass FUL it will pass those additional principles (if any) as well.[6] Thus, provided that a maxim passes FUL, it is indeed permissible to act upon it.

I think, therefore, that we can put aside the potential complications that threatened to arise from the existence of additional tests beyond that provided by FUL. We can say, straightforwardly, that if a maxim passes FUL then it is permissible to act on it. And we can combine this result with a point already made, that if a maxim *fails* FUL it is *forbidden* to act on it. Summing all of this up then we can conclude, quite simply, that it is permissible to act on a maxim if and only if it passes FUL.

It would, however, be easy to become confused about what we have shown so far. Suppose that in some situation I consider a maxim, M, that would permit me to perform an act, A, in those circumstances. And let us suppose, as well, that this maxim fails FUL. It would be natural to think that what this shows me is that it is forbidden to do A (at least, in these circumstances). But in point of fact this doesn't actually follow at all. From the mere fact that M fails FUL, all that immediately follows is that one should not act on M. That is, one should not do A for the particular *reasons* given by M. The maxim M, after all, corresponds to a particular reason-giving principle, and that principle picks out certain features of the situation, and tells me that by virtue of these features I have reason to do A. The fact that M fails FUL shows me that this particular claim about what I have reason to do (and why) is mistaken. Thus, if I *do* have reason to do A it is not for *those* (purported) reasons. But all of this is still compatible with the possibility that there may be other (genuine) reasons to do A — even in this very situation. For there may still be some other reason-giving principle which *is* sound — a principle that focuses on different features of the very same situation, and tells me that by virtue of *those* features I have reason to do A. In short, even though M fails FUL, some other maxim that would permit me to do A may still pass.

Thus, even though M fails FUL, we cannot yet conclude that it is forbidden to do A (in this situation). To reach that conclusion we would need to examine various other maxims as well, that is, the various other maxims that would also instruct me to do A. It is only if *all* such "permission giving" maxims fail FUL as well that we can conclude that doing A is forbidden. (Since I may only act on maxims that do pass FUL, if *all* such permission giving maxims fail, then I am indeed forbidden to do A.)

This is not to say, of course, that before concluding that a given type of act is forbidden in a certain situation one must literally examine a huge (perhaps infinite) number of maxims. It is possible that when a particular maxim fails FUL we will be able to see precisely why it fails, and generalize to other, relevant maxims. In effect, we may be able to test large classes of maxims at (more or less) the same time. But logically speaking the point remains, that the failure of a single maxim does not suffice to establish that a given act is forbidden; that requires, rather, the failure of all maxims that would permit the act (in those circumstances). (Similarly, of course, to establish that a given type of act was forbidden under *all* circumstances, we would need to show that all such permission giving maxims would fail, regardless of what circumstances the maxims specify as relevant.)

In the last few paragraphs I have been freely talking about actions as permissible or forbidden. What kind of permissibility is this? So far, the answer is *rational* permissibility. FUL provides a test for reason-giving principles, allowing us to conclude, in certain cases, that an action is rationally forbidden (say) because no genuinely adequate reason supports doing it. But the kantian believes that FUL captures a central *moral* idea as well. It serves to sort the morally permissible from the morally forbidden. If this is right, then rationality meets morality here: if the autonomy of reason requires you to conform to FUL, and acts forbidden by FUL are morally forbidden, then reason requires you to obey morality.

To understand why the kantian thinks FUL can plausibly be taken not only as a requirement of rationality but also as the basic principle of morality, it may be helpful to turn to a concrete example. Kant asks us to consider a case where I attempt to borrow some money, promising to pay it back, even though I know full well that I will be unable to keep such a promise. (The same basic example is discussed at two different places — G 4:402–3 and 422 — though only the second discussion makes explicit that the case involves money.) Kant supposes that my maxim here tells me that "when I am in a tight spot" I will "make a promise with the intention of not keeping it" (G 4:402). And here is part of Kant's discussion of whether this maxim passes FUL:

> I ask myself: would I be content with it if my maxim (of getting myself out of embarrassment through an untruthful promise) should be valid as a universal law (for myself as well as for others), and would I be able to say to myself that anyone may make an untruthful promise when he finds himself in embarrassment which he cannot get out of in any other way? Then I soon become aware that I can will the lie but not at all a universal law to lie. (G 4:403)

Several details of this argument will require more careful discussion later. Here I only want to draw attention to the plausibility of the idea that FUL is indeed concerned with fundamental moral aspects of the situation. In effect, Kant is telling us that immorality is a matter of cheating—making an exception of oneself (cf. G 4:424). When I tell a lie, or make a promise I don't intend to keep (or butt in line, or kill someone for personal gain, and so forth), I am playing by rules that I don't favor others acting on as well. After all, it is not as though someone who is immoral wants others to act in the same way! On the contrary, what I want when I act immorally is that everyone else should play by one set of rules (the moral rules) while I alone get to act on a *different* set of rules. Here I am, then, proposing to act in a certain way, in a certain situation, but it is perfectly clear that I cannot rationally will that everyone act in the same way in similar situations. There is a (purported) reason-giving principle that I propose to act on, but I can't reasonably favor that others act on it as well. This is the telltale sign of immorality, says the kantian. I want to treat myself differently than everyone else gets treated; I want one set of rules for myself, and another set of rules for everyone else. When I violate FUL, acting on a principle that I cannot will to be universal law, I try to make an exception of myself, even though I see full well that there is nothing at all that I consider a relevant difference between myself and others; and that is the mark of immorality.

That is why FUL is a requirement, not only of rationality, but of morality as well. And so we can conclude: if an act is forbidden by FUL, it is morally forbidden. But can we similarly conclude that if an act is permitted by FUL, it is morally permissible? As before, however, this conclusion assumes that FUL is not only one test among many, but is indeed the only fundamental principle—now, the only fundamental *moral* principle. This, too, is a claim that Kant appears to make (though it is not clearly distinguished from the earlier claim that FUL is the only fundamental rational principle), and for the sake of argument, at least, let us grant it as well (we'll consider its plausibility later). Then we can say that an act is morally permissible if and only if it is permitted by FUL.

Once again, it is important to avoid misunderstanding. We have just concluded that an act will be morally permissible if and only if it is permitted by FUL. And as we have already discussed, an action will be permitted by FUL provided that there is some maxim that passes FUL that permits the action in the circumstances. Note, however, that nothing that we have said requires that this maxim be the one that the person is actually acting upon. Provided that there is *some* permission giving maxim that passes FUL, it will be morally permissible to perform the act in question, even if the person is acting on some *other* maxim, and that maxim *fails* FUL!

Of course, if the person *is* acting on another maxim, and that maxim fails FUL, there will be plenty that is amiss. The person will be acting on a maxim that is unsound, both rationally and morally. That is to say, she will be performing the action for the *wrong* reasons—for "reasons" that are not actually adequate reasons for action at all. What's more, she will be performing the action for reasons that are not morally legitimate. As such, the person may well be open to moral condemnation of one sort or another. But this is not to say that what she is doing is morally *forbidden*. Rather, we will have a case of someone who is doing an action that is perfectly permissible morally, but is doing so for the wrong reason. In kantian jargon we can say that such a person is conforming to the moral law, but not acting for the sake of the moral law (G 4:390).

The distinction being drawn here is a perfectly familiar one. We all have the idea of someone doing the morally right thing, but for the wrong reasons. For example, Kant discusses a shopkeeper who gives correct change to his customers, but does so only out of fear of being caught and having business suffer (G 4:397). Presumably, we will all agree that giving correct change is a morally permissible (indeed morally obligatory) thing to do. And so we would agree that when the shopkeeper does this his action is morally permissible; he is conforming to the moral law. This is true even though he acts out of fear—acts for the morally wrong reasons. Thus, despite the fact that the maxim he acts on is unsound, that it would (as we may suppose) fail FUL, it remains true that the action he performs is morally permissible. And what *makes* it morally permissible is the very fact that some *other* maxim that enjoins giving correct change *would* pass FUL. In short, an act is morally permissible if and only if some permission giving maxim passes FUL, whether or not the person in question is actually acting on that maxim.

Let us now return to the question, earlier set aside, of how exactly we are to determine whether or not a given maxim does pass FUL. The basic idea, of course, is clear: a maxim passes FUL just in case I can will it to be a universal law. But how, exactly, can I tell whether or not I can "universalize" a maxim in this way? What, exactly, do I do when I try to determine whether a maxim can be universalized?

On what I take to be the standard proposal here, I should begin by trying to imagine a world where everyone *does* in fact conform to the reason-giving principle corresponding to the maxim being tested. If the maxim enjoins me to perform an act of type A, in such and such circumstances, then I am to imagine a world in which *everyone* performs acts of type A when in circumstances of that sort. I attempt to imagine a *full compliance world*, as we might call it, and then I ask myself two questions about this world. First, is such a world truly possible, or does something go wrong in trying to imagine it? Second, assuming that such a world is indeed possible (that nothing goes wrong in the relevant sense), can I rationally will it? The first question, in effect, is supposed to tell me whether the principle corresponding to the maxim could actually *be* a universal law; the second, whether I can *will* it to be such. To pass FUL, I must be able to answer both questions in the affirmative.

According to this interpretation, then, there are two distinct ways in which a maxim could fail to pass FUL, corresponding to the two questions I've just distinguished. In effect, there are two distinct subtests. This seems to be Kant's own view of the matter, in any event: he says that some maxims "cannot even be *thought* without contradiction as a universal law," while other maxims that also fail FUL generate no such "internal impossibility"; for these other maxims, rather, the *will* would "contradict itself" if it attempted to will the maxims to be universal laws (G 4:424).

More significantly, Kant seems to think this distinction picks out something important, generating different types of moral requirements. These are obscure matters, and Kant says little about them in the *Groundwork*, but roughly the picture seems to be this: when maxims fail at the first step, this is supposed to generate "perfect" duties, while "imperfect" duties (which are, despite the name, perfectly genuine duties) are generated by maxims failing at the second step (cf. G 4:424). But it is far from obvious why the two subtests should be invested with anything like this kind of significance. FUL says that one should not act on maxims that cannot be willed to be universal law. It does *not* say that it matters *why* a given maxim cannot be so willed. So it is far from clear that the kantian should follow Kant in holding it significant at which step a given maxim fails.

For that matter, it must be admitted as well that it is far from clear what precisely we are supposed to be concerned with as we consider the two subtests. In a moment we will turn to an examination of some of Kant's own examples. At the very least this should help us get clearer about what Kant thought could lead to a maxim's failing FUL. Whether, at the end of the day, we agree with Kant that it makes a difference at which step a maxim fails (or whether, indeed, maxims can fail in only two basic ways) is a matter of less importance.

Kant discusses four main examples in the *Groundwork*. I am going to discuss only two of these, but I am going to do so in some detail. (In thinking about these examples, it is also worth bearing in mind the point that Kant is only human. In certain cases he may simply be wrong about what FUL entails. Kantians can embrace FUL while still rejecting one or more of Kant's own views concerning which particular moral requirements emerge from it.)

A. The False Promise

The first example I want to examine is Kant's second, a return to the false promise case that we have already had a look at. Recall that Kant claims that "I can will the lie but not at all a universal law to lie." Here is Kant's initial argument for this claim — that I cannot will the maxim to be a universal law:

> for in accordance with such a law there would properly be no promises, because it would be pointless to avow my will in regard to my future actions to those who would not believe this avowal, or, if they rashly did so, who would pay me back in the same coin; hence my maxim, as soon as it were made into a universal law, would destroy itself. (G 4:403)

And here is the argument the second time around, when Kant returns to the case as one of his four examples:

> Yet I see right away that it [my maxim] could never be valid as a universal law of nature[7] and still agree with itself, but rather it would necessarily contradict itself. For the universality of a law that everyone who believes himself to be in distress could promise whatever occurred to him with the intention of not keeping it would make impossible the promise and the end one might have in making it, since no one would believe that anything has been promised him, but rather would laugh about every such utterance as vain pretense. (G 4:422)

Now the basic line of argument here is clear enough. If we try to imagine a world in which everyone lies, or makes insincere promises, so as to achieve personal goals by deceiving others, we find that something goes wrong. No one would believe you when you tried to make such a promise.

But what, exactly, is it that goes wrong here? Some think that what we find is that it is *literally impossible* for there to be a world in which everyone lies or makes insincere promises. Perhaps in a world where promises are so routinely broken, the very institution of promising would disappear (or, alternatively, would never have come into being). So there cannot be a world in which everyone makes promises they do not intend to keep. The maxim of lying to get out of a tight spot could not be a universal law, because there literally could not be a world in which everyone complies with this maxim. This interpretation sits nicely with Kant's saying that in such a world "there would properly be no promises," that the universality of the law would make such promises "impossible" — that (as he later puts it) the maxim "cannot even be *thought* without contradiction" to be a universal law (G 4:424). If there literally cannot *be* a world in which everyone acts on the maxim, I cannot will it to be universal law, and the maxim fails FUL.

Others interpret the argument somewhat differently. Taking their cue instead from Kant's remarks that making the promise in such a world would be "pointless," that it would be impossible to achieve "the end one might have" in making the promise, they conclude that what actually goes wrong is this: in a world in which promises are routinely broken, it is much more difficult, and perhaps even impossible, to achieve the *goal* specified in the maxim (getting out of a tight spot by deceiving others) by performing the *action* specified in the maxim (making an insincere promise). Insincere promising works more effectively (and perhaps only at all) against a general background in which people keep their promises. Thus, making the maxim be a universal law — one that everyone has reason to act on — undercuts the effectiveness of the maxim itself. And this involves a kind of practical contradiction: if I will my maxim to be universal law, I make it harder to achieve the very goal specified by the maxim by acting on that maxim. From the point of view of someone willing the maxim, then, it is irrational of me to will it to be universal law. So I cannot will the maxim to be universal law, and it fails FUL.[8]

This second interpretation, it should be noted, assumes that it is not rational for someone who accepts the maxim to *will* that everyone act on the maxim, since this makes it harder to achieve the goal specified in the maxim (in the specified manner). The argument thus presupposes some principle to the effect that it is not rational to favor things that make it harder to achieve one's goals. This is not an objection to the argument, of course, for presumably we would indeed want to endorse *some* such principle of instrumental reasoning (although the details of the principle might be a matter of debate). Kant himself, for example, earlier in the *Groundwork* (G 4:417), defends the claim that "Whoever wills the end, also wills (insofar as reason has decisive influence on his actions) the means," and it looks as though, on the second interpretation, this principle, or some near relative of it, is assumed.

Again, this observation is not intended as an objection to the second argument. If we are to sometimes reject maxims on the grounds that we cannot autonomously will them to be universal laws, then presumably one reason this may happen is because we find that willing the maxim to be universal law would be an act that would fall short in terms of one or another standard that we rationally endorse. Thus, if we do rationally endorse some principle of instrumental reasoning, it is not problematic for the kantian to appeal to that principle when arguing that one cannot will a particular maxim to be universal law.

Regardless of which interpretation we accept, it is worth drawing attention to the fact that the argument makes use of various contingent, empirical facts. The argument assumes, for example, that people have memories, and will recognize the fact that promise breaking has become widespread, and that this will result in either a breakdown of promising (on the first interpretation) or a disinclination to trust the promises of others (on the second interpretation). I note this point only to put to rest the widely held belief that kantians think that morality is entirely *a priori*, something that can be established without appeal to empirical facts.[9] At best, FUL itself has this kind of status. As we can see, however, more specific moral conclusions — such as a prohibition against lying or making insincere promises — are derived from FUL through the use of empirical truths.[10]

Suppose we grant, if only for the sake of argument, that Kant has successfully shown that the maxim in question cannot pass FUL. For reasons we have already discussed, however, it won't yet follow that it is morally forbidden to make an insincere promise in this case. To reach that conclusion, after all, we must argue that not only this maxim, but any other maxim that would permit lying here, would fail as well. Kant doesn't try to generalize his argument, to cover the other relevant maxims, but it is easy to see how the attempt might go. The features of the maxim that seem relevant to its failure are ones that would appear in any permission granting maxim relevant to the case at hand. Thus if one maxim that would permit lying here would fail, others should as well. (We'll consider an objection to this claim below.)

Can the argument be generalized even further? Can we derive not only a prohibition against lying in this particular case, but a general prohibition against all lying whatsoever? Kant thought so, and notoriously claimed that it is never morally permissible to tell a lie. But this is a point at which many kantians part company from Kant himself. To take the standard case, many kantians believe it permissible to lie to a would-be murderer so as to protect his innocent victim hiding in your basement. And it seems at least possible that á maxim that would permit this act could pass FUL. After all, the existence of at least some insincere promising is compatible with the continued existence and effectiveness of promising (for there are, let us admit, insincere promises made in the real world, yet promising has not been rendered impossible or ineffective). Thus it seems possible that a maxim that enjoined promise breaking or lying in sufficiently rare or special circumstances (for example) might yet pass FUL. Perhaps a maxim that permitted lying to the would-be murderer is one such.[11] Kant may have thought that FUL supported an absolute prohibition against lying, but the kantian need not follow him in this regard.

B. The Maxim of Nonaid to Others

Kant's fourth example involves a person who has a chance to aid another in need, but is tempted to pass him by without offering assistance. Kant imagines the person's maxim to be one of complete refusal to provide aid ("I will not take anything from him or even envy him; only I do not want to contribute to his welfare or to his assistance in distress"), and he says of this case:

> But although it is possible that a universal law of nature could well subsist in accordance with that maxim, yet it is impossible *to will* that such a principle should be valid without exception as a natural law. For a will that resolved on this would conflict with itself, since the case could sometimes arise in which he needs the love and sympathetic participation of others, and where, through such a natural law arising from his own will, he would rob himself of all the hope of assistance that he wishes for himself. (G 4:423)

Once again, the basic line of argument is fairly straightforward. Kant says that although there could be a world in which everyone acts on the maxim in question — a world where no one helps others — you cannot will this maxim to be universal law. You cannot rationally will that indifference to the needs of others be universal law, for you might find yourself in a situation where you *need* the help of others. The maxim thus fails at the second subtest: it cannot be willed to be universal law.

The first thing to notice about this argument is that it, too, appeals to empirical facts, here the fact that each of us has needs that we cannot always meet on our own. The second thing to notice is that it, too, makes use of something like the principle of instrumental reasoning. The thought is that each of us has goals of some sort, goals that we will want to achieve. But this makes it irrational to favor things that would make it more difficult to achieve those goals. Yet this is precisely what we will have done, in at least certain logically possible scenarios, if we will that it be *universal law* that no one help another. Once I recognize that I, too, can be in need of the aid of others, I cannot rationally favor a principle that would mean that I not get the help I need.

Notice, as well, that the relevant question is not particularly what I *would* will, *were* I in the situation where I needed help. That is no more relevant than the question of what I would will in the case where I don't need help. Rather, the question is what I am rationally prepared to will *here and now* as a principle to govern the case where I need help. Presumably, it is not rational of me (here and now) to be indifferent to my own need in that possible case. So I cannot (here and now) favor a rule that would mean that that need would go unmet (were it to arise). That is why I cannot will the maxim to be universal law. Thus it fails FUL.

Being clear about this point helps us to understand why it is irrelevant for someone to object that in the actual world they simply do not need anyone's help. Even if that were the case, it would remain a live possibility that the situation could be different: for anyone other than a deity, one *could* find oneself in need. And the thought, then, is that it cannot be rational to will, with regard to such a situation, that one not get the aid one would need.

Sometimes it is thought that whatever the force of this argument, it fails against an imagined "rugged individualist" who truly favors getting by completely on his own. Such a person, it is suggested, can will that the maxim of nonaid be universal law — for when he contemplates the possibility that he would himself be in need of the aid of others, he insists that even in such a case he (here and now) prefers that he die (in the given case) rather than be helped by others. (Of course, were he actually in a position of extreme need, he might lose his resolve and desire help. But as we have seen, that is strictly irrelevant. What matters is that here and now he wills that he not be aided, even in that case.)

I believe, however, that the kantian may have an answer to this objection available to him. For even the rugged individualist wants help of a particular kind—namely, to be left alone. This is easily seen if we imagine someone else bent on "aiding" him, despite his protests. The individualist wants the cooperation of others, just as the rest of us do; it is just that aid and cooperation take an unusual form in his case: leaving him to do things completely by himself. If this is right, then not even the individualist can favor a principle that would enjoin everyone to refuse to provide each with the particular aid that they need, for that would strip the individualist of what he most needs—to be left alone. If this is right, then none of us—not even the rugged individualist—can will a maxim of nonaid to be universal law.

Of course, as always, even if this is right it doesn't yet show us that it is morally forbidden to refuse to aid others. Doing that would require showing that not only this particular maxim but other, similar maxims would fail FUL as well. Once again, Kant doesn't attempt to generalize the argument, but here too it is not difficult to see how that more general argument might go: all humans (at least) are finite in ability, capable, in principle, of needing help (of some sort) from others; thus for any maxim at all that would simply permit disregarding the needs of others, no one can rationally will the maxim to be universal law.

But there remains a further worry. It might be objected that no principle at all could avoid the objection being raised against a principle of nonaid. For if, as the argument claims, it is irrational for me to will a principle (such as a principle of nonaid) that might leave me unable, or less able, to achieve my goals, then won't it be similarly irrational for me to will a principle that *requires* providing aid to others? After all, acting on a requirement to provide aid can itself leave me unable, or less able, to achieve one or another of my goals. Thus, won't the very same principle of instrumental reasoning that supposedly makes it irrational to favor a principle of nonaid also make it irrational to favor a principle *requiring* aid to others? How, then, can any principle at all—whether requiring aid or not—pass FUL?

Presumably the kantian must claim that an adequate answer to this worry involves balancing the various needs and aims I might have that might go unmet under the differing principles. I am looking for a principle that I can will to be universal law. And since, logically speaking, I might find myself in either one of the relevant roles (aid provider or aid recipient), I have to ask myself which costs I would rather endure. But in at least some cases—for example, when the gain to the needy when aid is provided is significantly greater than the loss to the person who actually provides the aid—the answer to this question is clear. Presumably, then, the principle of instrumental reasoning can lead me (here and now) to favor principles that do require providing aid in cases of this sort. But if this is right, then FUL will indeed support some sort of requirement to provide aid after all.

Doubtless, further questions could be raised about both of these examples (and as I have already noted, Kant discusses two other examples in the *Groundwork* as well). But I hope I have said enough to give at least some sense of how FUL is supposed to be used as a test for maxims and for deriving moral obligations.

Our discussion should also put to rest one common objection to FUL, namely, that it has no "bite," that any maxim at all can pass. For as we have now seen, it's not implausible to think that certain maxims do indeed fail FUL. Thus, whatever its other shortcomings may be, at least FUL isn't altogether devoid of content.

There are, however, other general objections to FUL that merit further discussion. Let me quickly mention four. All of them concern the adequacy of FUL from the *moral* point of view. First, it is sometimes objected that FUL is raising a morally irrelevant concern when it asks us to consider a world where everyone acts on the maxim in question. After all (the objection notes), in the *real* world typically it simply isn't going to happen that *everyone* acts on a given maxim. From the moral point of view, then, why should we concern ourselves with such an unrealistic possibility?

Recall, however, that the kantian's position is that if a maxim passes FUL, then it is morally permissible to act on it, indeed, morally permissible for anyone at all to act on it. It hardly seems irrelevant, then, to consider a world in which everyone *does* act on the given maxim. This would simply be a world in which—in the relevant way at least—everyone is acting in a manner that is supposedly morally permissible. Surely it makes sense to insist that it must at least be *possible* for everyone to act in a morally permissible manner, and indeed, to insist further that it must be reasonable to *favor* a world in which everyone acts in a morally permissible manner. (It cannot be preferable, from the moral point of view, that some act in a morally forbidden manner.) A world in which everyone acts morally must be both possible and attractive. Thus, in directing our attention to a full compliance world, FUL is not at all directing our attention to a morally irrelevant possibility.

But this immediately suggests a second objection: even if the full compliance world is indeed a world worth considering when testing maxims from the moral point of view, it is quite another matter to suggest that this is the *only* world worth considering, or even the most important. After all, in the real world not everyone is going to act morally, and so it is important to know how one is permitted (or required) to act in the face of immoral behavior by others. It would seem that the relevant question with regard to such cases of *partial* compliance is what I can will with regard to a world in which *not* everyone is acting on the maxim in question. But FUL apparently never asks us to consider such worlds: it *restricts* our attention to asking whether I can will a given maxim in a world in which everyone is acting on the maxim. Thus FUL inappropriately disregards the very real possibility of immoral behavior (partial compliance). Worse still, because of this neglect, it can generate morally implausible guidance, since acts that might be perfectly attractive were everyone to be acting morally (ones that I can will for the full compliance world) might be catastrophic when done in the face of immoral behavior.

Presumably, this difficulty about how to properly evaluate maxims for dealing with partial compliance might not be particularly worrisome if there were further tests, beyond FUL, that needed to be passed as well before it was permissible to act on a given maxim. If there were such further tests, then they might do a better job of evaluating whether a maxim can properly handle cases of merely partial compliance. We could appeal to these further tests to rule out maxims passed by FUL that were inadequate in this regard. But as we have already noted, Kant believes that FUL (and its equivalent, alternative formulations) is the only fundamental principle needed, and kantians have typically followed him in this. So it is worth asking whether FUL has the ability to handle the problem of imperfect compliance on its own.

I believe that it does. The problem, I think, lies not with FUL itself, but with what I earlier called the "standard proposal" for interpreting FUL. According to this interpretation, recall, to see whether a maxim passes FUL I need only ask whether I can will that the principle corresponding to the maxim be one that everyone acts upon. That is, I need only consider the full compliance world—whether it is possible, and whether I can rationally favor it. But why should we take FUL to be so easily satisfied? According to FUL, after all, I should only act on maxims that I can will to be *universal* law. In particular, then, I have to ask whether the appropriate principle is one that I can rationally will for *all* cases to which it applies. Now one such case, to be sure, may well be the case of full compliance. But often enough the principle in question will apply to other cases as well, cases of imperfect compliance; and so I must ask whether I can rationally will that the principle govern *those* cases as well. Thus, contrary to the claim put forward by the objection (and reinforced by the standard interpretation), FUL does not actually disregard consideration of partial compliance worlds, worlds where not everyone is acting morally. On the contrary, it demands that we consider such worlds as well, before signing off on a principle. Only if we can will the principle for cases of imperfect compliance as well (assuming that it applies to such cases) is it really true that we can will the maxim to be universal law.

It may also be worth recalling, in this regard, that the principles we favor need not prescribe the same type of action regardless of circumstances. In particular, then, we might favor principles that tell us to act in one way when others are acting similarly, and in quite another way when they are not. Thus the principles that pass FUL may enjoin one kind of behavior when others are acting morally, and quite another in the face of immoral behavior. In short, there is no good reason to believe that FUL will be unable to generate appropriate moral guidance for dealing with cases of noncompliance.

A third objection complains that in point of fact *no* maxims (or perhaps only very few maxims) can actually pass FUL. In particular, perfectly harmless maxims—maxims that intuitively it ought to be permissible to act upon—fail. If this is correct, of course, then we have some reason to reject FUL: if it fails maxims that ought to pass, then it isn't a very good test of the validity of a maxim. Here is an example of the sort of problem that people have in mind when they raise this worry. Suppose that I form the intention of going to the local pizza house, and ask whether my maxim ("I will go to Naples for lunch") can pass FUL. I must ask whether I can will this maxim to be universal law; and apparently this involves trying to imagine a world in which *everyone*—at a minimum, all five billion humans—goes to Naples for lunch! But as soon as I do this I see that either this is literally not possible (not everyone could fit) or it would involve a practical contradiction (it would make it much more difficult to get lunch). Thus my maxim fails FUL. But this—the objection concludes—is absurd. Surely going to the local pizza house is morally permissible (special circumstances aside), and if FUL condemns my maxim, so much the worse for FUL.

In answering this objection, the first thing to remember is that even if this maxim does fail FUL, that doesn't entail that it is morally impermissible to have lunch at Naples. So long as another maxim that permits having lunch at Naples passes FUL, then it will be perfectly permissible to have lunch there. At worst, all that would follow is that the short maxim we are here testing—"I will go to Naples for lunch"—does not provide a completely accurate account of what I have reason to do. And this is not, in fact, an implausible claim. For as a moment's reflection makes clear, whether it makes sense for me to go to Naples depends on any number of factors not mentioned in the maxim as stated, for example, whether or not I am hungry, whether or not I want pizza, whether or not the restaurant is crowded, whether or not it is nearby, and so forth. Presumably I do *not* have reason to go to Naples regardless of how crowded it is, how inconvenient it is to get to it, and so on. Thus the simple maxim "I will go to Naples for lunch" cannot in fact be plausibly taken to be a complete account of what I have reason to do and why. That requires a much fuller statement, one that, for obvious reasons, I rarely have occasion to try to articulate fully. Normally, the relevant extra conditions are left implicit, and so the short maxim is perhaps best understood as a kind of shorthand for that fuller statement.

Once we keep this point in mind, and try to universalize an appropriately full statement of the maxim (or universalize the short maxim, understood to implicitly contain the various necessary qualifications), we find that the maxim can indeed pass FUL. I can certainly will that everyone go to Naples if it is convenient, if it isn't too crowded, if they want pizza, and so forth. After all, obviously enough, one or another of these conditions won't be met for almost any person we might consider (most, for example, are much too far away for it to be convenient). And so, when we imagine a world in which everyone acts on this maxim, we won't imagine a world with billions trying to crowd into the local restaurant. Rather, we imagine a world in which those who want pizza and are nearby (and so forth) go. And this is a world, it appears, that we can readily will.

In short, if we take the simple maxim to be a complete statement, it does fail FUL, but appropriately so, while the fuller maxim passes. And if we take the simple maxim to be shorthand for that fuller maxim, then of course it passes as well. Either way, there will indeed be a maxim that passes FUL that permits me to go to Naples (special circumstances aside), and so, contrary to the claim of the objection, FUL won't forbid this morally innocuous act.

The third objection claimed (albeit incorrectly) that too little passes FUL. The final objection that I want to consider, our fourth, makes the opposite complaint, that too much passes. For as we have just seen, a complete specification of one's maxim might include any number of clauses and conditions. (FUL does not restrict us to testing "simple" maxims: any maxim can be put forward for testing.) The worry, then, is that if one is sufficiently clever in formulating one's maxim, one can always arrive at a version that will pass FUL, no matter how morally unacceptable the act in question. For example, suppose I want to murder you. Even if (as we might suppose) the straightforward maxim "I will murder those I want dead" would fail FUL, I need only propose, instead, a maxim that includes, say, my proper name. Suppose, then, that I try the maxim "If I am named Shelly Kagan then I will murder those I want dead." If this maxim can indeed pass FUL, then I am permitted to murder you (whether or not this is in fact my maxim). But this would clearly be unacceptable. So if the rigged maxim does indeed pass FUL, we will simply have to reject FUL.

The objection then continues by insisting that this maxim does, in fact, pass FUL. After all, there is presumably no impossibility about having a world in which *everyone* named Shelly Kagan kills at will (indeed I may well be the only person named Shelly Kagan in the world), and it certainly seems that I (Shelly Kagan!) can be in favor of a principle that gives me this extra freedom. So it looks as though I can will the maxim to be universal law, and FUL unacceptably permits me to kill at will. (Similar results could presumably be achieved by replacing my name with a definite description that uniquely picks me out, for example, "If I am a professor of philosophy at a midsize university, with three children, and a wife who works as a midwife, etc., etc., . . . then . . ." For simplicity, however, I'll stick to introducing the proper name.)

In fact, however, I think it far from obvious that I can rationally will the maxim in question to be universal law. After all, although I believe that I am one of at best a handful of people named Shelly Kagan — perhaps, indeed, the only one — I could presumably be mistaken about this. Perhaps there is a vast extended clan, currently living peacefully in the jungle, all of whose members are named Shelly Kagan. I can hardly rationally favor a principle that would permit this vast group to kill at will. And even if (as certainly seems likely) this possibility is unrealized in the actual world, there *could* be such a world, and it simply isn't true that I (here and now) am prepared to will with regard to such a world that all the Shelly Kagans in that world be permitted to kill at will. Thus it isn't really true that I can rationally will that the maxim "If I am named Shelly Kagan then I will murder those I want dead" be a universal law. Accordingly, the fourth objection fails as well.

Generalizing from the failure of this particular example, it seems we can say the following. Although nothing in FUL, in and of itself, places restrictions on the content of the maxims that we bring for testing — we can add whatever silly clauses and conditions we'd like — proper application of FUL does have the result of ruling out maxims that introduce irrelevant conditions. If a maxim is couched in terms of conditions that are in point of fact rationally and morally irrelevant, we will discover that we are not genuinely prepared to will that the maxim be a universal law.

But the discussion of the third objection has already suggested a complementary point as well, namely, that proper application of FUL will also have the result of ruling out maxims that *lack* relevant conditions. If a maxim is overly simplistic, we will find that we are not genuinely prepared to will that either. Taking these points together, then, the kantian claims that FUL provides a sufficiently subtle and sophisticated test to guide us toward plausible moral principles, ones that are sensitive to the relevant features of acts and their circumstances while disregarding the irrelevant features.

III. Kantianism and Consequentialism

What would those moral principles look like? That is, given FUL, what kind of normative moral theory emerges? Putting the question like this appropriately emphasizes the fact that what we have been primarily discussing up to this point is the kantian account of the *foundations* of ethics. (In this regard we have been following the lead of Kant himself in the *Groundwork*, the very title of which, after all, reveals that its primary concern lies with foundational issues.) We have not yet much concerned ourselves with describing the particular *normative* principles (roughly, the more directly action guiding principles, such as those requiring promise keeping or aiding others) that would emerge from that account, except as a means of illustrating FUL at work.[12] I have, of course, tried to portray that kantian account of the foundations of ethics as attractive and worth taking seriously—and if I have succeeded in this endeavor, then my primary purpose in this essay is accomplished. Still, it is natural to wonder about the normative level as well. Given FUL, what kinds of normative principles are we led to?

We have, of course, already taken a quick look at two particular examples. FUL, we have seen, rules out moral principles that would permit me to be indifferent to the needs of others, or to lie (or make a false promise) simply because this would be personally convenient. Obviously enough, given the time, we could apply FUL to a variety of other cases as well, and doing this over a sufficiently wide range of cases would doubtless enhance our understanding of FUL's plausibility and adequacy. But instead of continuing to focus on particular cases, I want to step back and ask, in a general way, whether we can say anything helpful about the overall structure of the moral theory that would emerge from FUL.

I raise this question, of course, because most kantians have thought it fairly clear that FUL supports a deontological moral theory. Kant himself certainly believed this. Indeed, even those who reject deontology—consequentialists being the most prominent among this group—have typically accepted this claim as well, and thus concluded that avoiding deontology requires rejecting the kantian account of the foundations of ethics. Now it is certainly true that nothing that I have said in this essay constitutes a full defense of kantianism. One might reject the account of autonomy that I sketched at the outset, for example, or deny that autonomy leads to FUL. If one does this, of course, then even if it is true that FUL does support deontology, given a rejection of FUL this won't threaten, say, one's acceptance of consequentialism. On the other hand, some will find the kantian account of autonomy and its implications sufficiently attractive, and FUL sufficiently plausible in its own right, that they are prepared to accept the moral principles supported by FUL, even if this requires revising some of their previously held moral opinions. If FUL does indeed support deontology rather than consequentialism, this may then provide a powerful argument in favor of deontology.

But there is a third possibility as well, of course, which is that Kant and most kantians are wrong when they claim that FUL supports deontology. If it should turn out that FUL actually supports consequentialism instead, then to the extent that one finds the kantian account of the foundations of ethics attractive, this will actually provide an argument in favor of consequentialism, rather than deontology.

Of course, one point is certainly true. If the kantian account of the foundations of ethics is correct, then the *basis* of ethics looks rather unlike the accounts typically offered by consequentialists. For historically speaking, at least, most consequentialists (though certainly not all) have grounded their consequentialism in what we might call *foundational consequentialism* —the claim that the ultimate basis of the (valid) normative moral principles lies in an appeal to the significance of the overall good. In contrast, the kantian account that we have been sketching gives no particularly important role at the foundational level to the concept of the good at all. The ultimate basis of morality, for the kantian, is not the good, but rather freedom. For this reason, it is appropriate to say that the kantian account of the foundations of morality is foundationally deontological, rather than foundationally consequentialist.

But it is one thing to insist that the kantian account of the foundations of ethics is usefully classified as deontological; it is quite another to insist that the particular normative principles that emerge from that account are themselves deontological. For absent further argument, there is no particular reason to assume that deontological foundations must yield deontological moral principles.[13] When I claim, then, that FUL may well support consequentialism rather than deontology, I have in mind a claim not about the foundational level, but rather one about the normative level, the level that concerns the various action guiding principles themselves. FUL may itself be grounded in a nonconsequentialist account (this much certainly seems to be true), but what *emerges* from FUL may well be a consequentialist rather than a deontological normative theory.

Evaluating this claim, of course, requires at least a working account of the distinction between deontological and consequentialist theories (at the normative level). Simplifying somewhat, the following should do for our purposes. Consequentialism holds that an act is morally permissible if and only if it has the best overall consequences (of those acts available to the agent). Deontology rejects this simple account of right and wrong, insisting that certain acts are morally forbidden, even when they would lead to better results overall. Deontologists thus embrace *constraints* — prohibitions against performing the offensive types of acts, even when doing so would lead to better results. Typical examples of constraints include prohibitions against lying, harming the innocent, failure to keep one's promises, and so on.[14]

Deontologists normally also reject consequentialism on the further ground that it is too demanding, always requiring the agent to perform the act that would lead to the best results overall, no matter how great the sacrifice involved to the agent himself. Deontologists thus typically embrace *options* as well — permissions to avoid promoting the overall good when the cost to the agent would be too great. For example, deontologists typically don't believe we are required to sacrifice huge portions of our income to famine relief, even though if we did so a great many lives might be saved. Such sacrifice is doubtless praiseworthy (they say), but it is strictly optional: we are permitted, instead, to pursue our own individual projects — as well as going to concerts, eating at expensive restaurants, and so forth — even though our time and money could do much more good were it spent in other ways. Most deontologists do insist, of course, that sometimes sacrifices for others are morally required (for example, when I can rescue someone at minimal cost to myself); but consequentialism goes too far (they say) in putting no limits on the obligation to promote the overall good.

While most deontologists accept both constraints *and* options (thus holding that consequentialism sometimes permits what is actually forbidden, and sometimes requires what is actually optional), I think it fair to say that so long as a theory contains constraints, it would normally be considered deontological, whether or not it contained *options* as well. In contrast, the presence of options alone (that is, without constraints as well) would not suffice to render a theory deontological. For our purposes, then, in asking

whether FUL supports deontology, the key question facing us is whether or not FUL supports constraints.

Nonetheless, it may be helpful if we begin with the question of whether FUL supports *options*. For even though deontologists need not accept options, all consequentialists reject them. Thus, if FUL is to generate a consequentialist normative theory, it must reject options as well. Let us therefore postpone, for the moment, the question of whether FUL supports constraints. Even if there *are* constraints, I might still be morally required to do as much good as I can by permissible means (that is, those means not forbidden by constraints). So in asking whether FUL supports options or not, we are asking whether FUL supports a requirement to do as much good as one can — *within* the limits of constraints (if any).

Now we already know, from the discussion of the aid example, that FUL generates a requirement to aid others; FUL does not allow us to be indifferent to the good that we can do. But many kantians have thought it plain that FUL does not require us to do as *much* good as we can (within constraints — a qualification that I will hereafter leave implicit). While FUL sometimes requires us to promote the good (such as helping to meet the needs of others), it does not require us to do *all* that we can in this regard. The claim of these kantians, then, is that FUL generates a requirement to aid, but a *limited* one; when the cost of providing aid to others is too great, I am not required to do it.

But it is far from obvious that FUL will actually support this kind of limitation on the requirement to provide aid. It is certainly true, of course, that a maximally demanding requirement to promote the good will potentially impose considerable costs upon me. Indeed, in the real world I might find myself required to make huge sacrifices, while benefiting little, or not at all, from the fact that others are similarly required to promote the overall good. But in evaluating alternative principles concerning aid I must bear in mind the fact that I am looking for a principle that I can rationally favor for *all* worlds to which it applies. I cannot restrict my attention to the costs and benefits that I actually expect; I must consider all possible costs and all possible benefits. And since I have no more reason to be concerned with the costs that I might have to pay (as benefactor) than with the benefits I might receive (as recipient), it seems reasonable to favor a principle that provides the best overall balance of costs and benefits. But this is precisely what is done by a requirement to promote the *overall good*: it requires sacrifices only in those cases in which an even greater amount of good overall is thereby achieved. Thus, when I ask myself what sort of requirement to provide aid I can rationally favor to be universal law, it may well be that I must favor a requirement to bring about the best possible results overall. Anything less demanding will be inadequate.

Indeed, this implication of the aid example may have been staring us in the face all along, even if we did not previously draw it. For any requirement to provide aid at all will impose costs on those who have to provide the aid. If, nonetheless, I cannot rationally favor a maxim that would allow me to remain indifferent to the needs of others — and this, after all, is what Kant and kantians have always claimed — this must be because when I bear in mind the logical possibility that I might be either benefactor or recipient, I am led to balance the potential costs and benefits, and thus come to favor a principle that *at a minimum* requires aid when the benefits to those in need are significantly greater than the costs to those providing the aid. This is what we argued when discussing the original aid example. But this line of thought, if it is sound at all, has no obvious stopping point short of a general requirement to promote the overall good. The same balancing that leads me to favor a principle requiring aid when the benefits are "significantly" greater than the costs will, it seems, similarly lead me to favor a principle requiring aid *whenever* the benefits are greater than the costs, period. Thus, if FUL supports any requirement to provide aid at all, it should support a requirement to promote the *overall* good.

As always, there are a variety of objections that might be raised against the argument I have just been sketching. But once again, my purpose is not to offer a full defense of the claim that FUL rejects options. I merely wanted to indicate one main line of thought that might lead one to hold that FUL supports a general requirement to promote the overall good — despite what many kantians seem to believe.[15]

So let us suppose, if only for the sake of argument, that FUL does rule out options. As we have already noted, this is still compatible with FUL generating a deontological system. For we have not yet considered the question of whether FUL supports *constraints*. If it does, of course, then despite the general requirement to do as much good as possible within the *limits* of those constraints, it will still be true that certain kinds of action will be forbidden even when performing acts of the given kinds would lead to better results overall. Thus, so long as FUL supports constraints — even if it does reject options — it will in fact generate deontology rather than consequentialism. Accordingly, our next question must be whether FUL supports constraints.

Now it might seem obvious, in light of our earlier discussion, that FUL does indeed support constraints. For our very first illustration of FUL at work seemed to show that it rules out making insincere promises, or, more generally, lying. But if FUL does support a moral prohibition against lying, doesn't it follow trivially that it supports constraints, and thus that it supports deontology?

In point of fact, however, this conclusion does not follow so readily, for consequentialists themselves will be among those who support a moral prohibition against lying. Normally, after all, lying leads to worse results overall (counting everyone's interests equally) and so lying will typically be forbidden — even by consequentialists. In particular, in a typical case of false promising the overall results would be better if one refrained from making the insincere promise. Consequentialists will thus join deontologists in forbidding me to make insincere promises on the mere grounds that I need the money, or am in a tight spot, and so forth. And this means, of course, that from the mere fact that FUL prohibits making the insincere promise in such a case, we cannot yet determine whether FUL supports a *constraint* against lying and making insincere promises — even when (unlike the normal case) lying would have better results overall. Thus we are not yet in a position to tell whether FUL supports deontology or consequentialism. (Similarly, of course, for normal cases of promise breaking, harming the innocent, and so forth.)

What is needed, rather, if we are to settle the matter, is a case where it is stipulated that lying would lead to *better* results overall. If FUL would forbid lying even in a case of this kind, then indeed it would be clear that FUL generates a deontological normative theory — since it would support a moral principle that forbids lying even when lying is necessary to achieve the best results overall. But we have not yet investigated whether FUL prohibits lying even in cases of *this* sort; and I don't think it obvious that it does.

Of course, as we have already noted, Kant himself believed that FUL (or its equivalent) rules out *all* cases of lying, no matter what the circumstances. Were he right about this, obviously enough, FUL would be the basis of a particularly strict form of deontology. But as we have also noted, many kantians refuse to follow Kant on this matter, holding that under the *right* circumstances FUL can indeed pass a maxim that would permit lying (for example, lying to a would-be murderer). So at a minimum, we shouldn't take it as *obvious* that FUL will pass no maxims that permit lying when this is necessary to promote the overall good.

Presumably, we might attempt to settle the matter by considering a particular case where lying is stipulated to lead to better results overall, and then testing various maxims that would permit lying in such a case — so as to see whether any of these lie permitting maxims could pass FUL. In principle an investigation of this sort could tell us whether FUL forbids lying even when such an act leads to better results overall. If it does, this would show that FUL supports a constraint against lying, and thus supports deontology rather than consequentialism.

But such an investigation would have a variety of drawbacks. First of all, suppose we took some such maxim—say, a maxim of the form "I will lie under such and such circumstances"—and found that it could not pass FUL. As we know, this would show that one should not act on that maxim. But it would not actually show that FUL *forbids lying* in such cases. It would only show that one should not act on *that* maxim, that *if* lying is permitted, the reasons why it is permitted are not adequately captured in the particular maxim being tested. It would still be possible that some other maxim would pass FUL, a maxim that would permit lying in the case at hand.

On the other hand, suppose we found a maxim that permitted lying in the particular case imagined. That would of course show that it was permissible to act on that maxim, and thus permissible to tell a lie in that case, and thus—by hypothesis—permissible to tell a lie in at least one case where doing so leads to better results. But this still wouldn't necessarily constitute a defense of consequentialism, for the fact that lying here leads to better results might be irrelevant (or inadequate, by itself) to explaining *why* the maxim passed FUL. There might well be other cases where lying would also lead to better results, yet where telling a lie would *not* be permitted by the particular maxim that permitted it in the original case. In short, even if lying is permitted in *some* cases where this happens to have the best results, we couldn't necessarily conclude that it was permitted in *all* cases where this had the best results. So even if we did find a maxim that permitted lying in our original case, we wouldn't necessarily have shown that FUL supports consequentialism. To do that, we will need to show that FUL permits lying in *all* cases where this has the best results overall.

But of course even this wouldn't suffice, for it might be that FUL permits *lying* in all cases where this leads to the best results but nonetheless rules out *other* types of actions, regardless of the results. Perhaps, for example, lying is permitted when this promotes the overall good, but there is, nonetheless, a constraint against bodily harm to the innocent, even when this is necessary to bring about the best results overall. If something like this were the case, then, of course, it would still be true that FUL supports deontology. So long as there is any constraint at all—any prohibition against performing an act with good results overall—FUL supports deontology rather than consequentialism. In short, focusing on maxims concerned with lying alone will be too narrow a method of investigation to settle the question of whether FUL supports deontology or consequentialism.

What we want to know, of course, is whether there are any actions at all, of any sort whatsoever, that are forbidden even when performing actions of that sort is necessary to bring about the greatest amount of good overall. If any act, of any kind, is forbidden even when the results would be better, then FUL supports deontology. What the consequentialist must insist, therefore, is that *any* act is permissible, so long as it leads to the best results overall. But since the permissibility of an act follows so long as there is a single maxim that passes FUL that permits the given act, what the consequentialist must claim is that for each act that has the best results, there is *some* maxim or the other that would permit the act, that passes FUL.

Now in principle, I suppose, it could be a different maxim in each case. But this hardly seems likely. For as we have seen, maxims that pass FUL are supposed to do so by virtue of referring to the various features of the situation that actually provide the agent with adequate reason for acting in the specified manner. A valid, fully specified maxim would pick out all and only those features of the situation that make it reasonable for the agent to act in the given way. According to consequentialism, however, what *ultimately* justifies an agent's performing a given act is always the very *same* reason, namely, that the act would lead to the best results overall. Thus the consequentialist believes that in any given case, the act that leads to the best results is the appropriate act to perform, and the ultimate *reason* why it is the right act to perform is the very fact that it leads to the best results. Thus we should expect the consequentialist to hold that the principle "act in the way that has the best results overall" is universally valid, and that the quite general maxim "I will act in the way that has the best results overall" will pass FUL (no matter what the particular case at hand).

So let us consider that maxim. If it passes FUL, then, of course, it is permissible to act on it, which means that it will always be permissible to perform the act that has the best results overall—*whatever* type of act that may be, and whatever the circumstances. In short, if the *consequentialist maxim* (as we might call it) passes FUL, then it is never forbidden to perform the act with the best results, FUL does not support constraints, and FUL does not support deontology.

Does the consequentialist maxim pass FUL? I believe it does. At the very least it must be admitted that if it fails FUL it is not obvious how and why it does so. Consider the sorts of difficulties that have plagued maxims in our previous examples. On at least one interpretation of the false promising example we literally cannot imagine a world in which everyone makes false promises. Is there any comparable impossibility with regard to a world where everyone acts in such a way as to produce the best results overall? Obviously not. A world where everyone promotes the overall good is, sad to say, highly unrealistic, but there is no conceptual impossibility involved in trying to imagine it.

On the alternative interpretation of the false promising example, the existence of a world where everyone makes false promises makes it more difficult to achieve the end specified in the maxim itself through the means specified by the maxim (that is, getting out of a tight spot by making a false promise is less likely in a world where everyone tries to do this). When one imagined the maxim as universal law, the maxim's course of action became a less effective means to the maxim's own end. This was a kind of practical contradiction. Is there any comparable practical contradiction involved in imagining a world where everyone promotes the overall good? Again, the answer is obviously not. A world where everyone promotes the overall good is not a world that makes it more difficult to bring about the best results overall. On the contrary, it is likely to be a world that makes it easier to bring about the best results overall. Thus, whatever our interpretation of the first step of the FUL test, there seems to be no reason to think that universalizing the consequentialist maxim leads to a "contradiction in thought."

Nor, so far as I can see, is there any reason to believe that universalizing the consequentialist maxim leads to problems at the second step, generating a "contradiction in will." When we imagined a world where no one aided others in need, this was indeed a coherent possibility, but we found we could not will the relevant maxim to be a universal law. Given that we ourselves could have needs (that we were unable to meet without aid from others), it violated a principle of instrumental reasoning to favor a maxim that if made a universal law would necessarily leave those needs unmet. But is there any comparable violation of instrumental reason involved with willing it to be a universal law that everyone is to bring about as much good as possible? It is far from clear that there is.

To be sure, if it is a universal law that everyone is to bring about as much good as possible, then there may arise cases in which I may have to make significant sacrifices for others. From the point of view of instrumental reasoning this is undesirable, and gives me some reason to oppose such a requirement. But we have, of course, already considered this point. Since I am asking what I can will to be universal law, I must also consider the possibility that I might be the *recipient* of the aid. In effect, I must weigh all the potential costs against all the potential benefits, and when I do this—or so I have argued—instrumental reasoning will lead me to favor a principle in which sacrifices are required precisely when the benefits are greater than the costs. That is to say, instrumental reasoning will lead me to favor a principle requiring each of us to act in the way that has the best results *overall.*

Are there other reasons to think that I cannot rationally favor its being a universal law that everyone is to act in such a way as to maximize the overall good? At the very least it is not obvious what they might be.

Of course, one might object to such a law on the very ground that it would permit violating *constraints!* Intuitively, after all, certain acts are simply morally forbidden, despite their results. But promoting the overall good might sometimes require performing acts of these intuitively unacceptable kinds. Isn't this adequate grounds for refusing to will the maxim to be a universal law?

In point of fact, however, it is not at all clear that such intuitions are even *relevant* in thinking about which maxims pass FUL. FUL, after all, was supposed to be the basis of morality, the source of the valid moral principles (whatever they turn out to be). It can't play this role if we are going to *presuppose* various moral principles (whether directly, or by relying on moral intuitions) in determining what can, and what cannot, pass FUL. Put another way, given the kantian account of the foundations of ethics, appeals to moral intuitions are logically beside the point, until we have confirmed their accuracy *independently*, through appeal to FUL (cf. G 4:408–10). Thus we cannot appeal to the intuitive plausibility of constraints, and use this as a reason for claiming that principles that violate such constraints must fail FUL. Rather, we must first decide what passes FUL—and we must do this on independent grounds. And what this means, of course, is that despite the intuitive appeal of constraints, we don't yet have reason to think that FUL *generates* constraints.

For all that, of course, there might well be further arguments available to those who want to claim that I cannot rationally will it to be a universal law that everyone do the act with the best results overall. If such further considerations were offered, and found to be compelling, then it would indeed turn out that the consequentialist maxim cannot pass FUL. I certainly haven't attempted to discuss all possible arguments along these lines. But it must be admitted, I think, that it isn't obvious what these further arguments might look like. And so I think we should conclude—even if only tentatively—that our maxim can indeed pass FUL. Or, at a minimum, we should at least admit that this possibility is not one that can be readily dismissed.

But if the maxim passes FUL then it is always permissible to act on it. It is always permissible to do the act that will have the best results overall. Thus, if the consequentialist maxim passes, there are no constraints. FUL simply doesn't generate them.

Putting together the results of these various arguments, we can say, at a minimum, that it should not be taken to be obvious that FUL supports a deontological normative theory. On the contrary, there is at least some reason to believe that FUL yields no constraints at all, despite what Kant and most kantians have assumed. Indeed, there is some reason to believe that FUL supports a normative theory with neither constraints nor options. On such a theory, each of us is simply required to do as much good as possible. But this, of course, is consequentialism.

Here is a slightly different way to see how consequentialism is supported by FUL (assuming that the arguments we have been considering are sound). We have just argued that despite what kantians have typically thought, it may well be the case that kantian foundations support the claim that it is always *permissible* to do the act that will have the best results overall. By itself, of course, this result (even if correct) wouldn't yet show that we are *required* to do the act with the best results. But this further conclusion would indeed follow given the earlier claim that we are required to do as much good as possible within the limits of whatever constraints there may be. For if we are always *permitted* to do the act with the best results, there *are* no constraints. Thus the requirement to do as much good as possible *within* the limits of constraints reduces to the simple requirement to do as much good as possible. Each of us is required to do the act with the best results overall. But this, again, is precisely the claim of consequentialism.

These same basic ideas (if they are accepted) can be rearranged once more, into an even more straightforward "proof" that FUL supports consequentialism. To begin with, since the consequentialist maxim passes FUL, agents are always *permitted* to perform the act with the best results overall. But in point of fact, contrary to what most people have thought, no *other* maxim will pass FUL as well[16] (since any maxim that permitted doing less, or required doing something different, would run afoul of the principle of instrumental reason, and thus could not be willed to be universal law). Thus agents are actually *required* to do the act that would best promote the overall good. In short, given FUL—and assuming, of course, that the arguments we have been considering are correct—everyone is required to do the act with the best results overall, just as consequentialism claims.

Once again, it is worth emphasizing that I do not take these remarks to constitute a full defense of the claim that kantian foundations support a consequentialist normative theory rather than a deontological one.[17] But I hope I have said enough to show that this possibility is one that must be taken very seriously indeed, despite the fact that Kant and almost all kantians after him have rejected it (as have indeed almost all those who have studied kantianism, whether sympathetic to it or not).

If one were to attempt to complete the project of grounding consequentialism on a kantian basis, much would still need to be done. Beyond the obvious point, that the various arguments sketched here would need to be developed more fully (and a host of objections would need to be considered in greater detail), the most important remaining task would be this. A consequentialist theory is incomplete until combined with a theory of the good. Knowing that we are required to do as much good as possible does not yet generate determinate guidance until we know what makes one outcome better or worse than another. What we need, then, is an account of the intrinsic goods for the sake of which we should act. If the kantian account of the foundations of morality is correct, of course, then the intrinsic goods must be ones that we can autonomously set for ourselves as ends. Kantians believe there are such goods, however, and so the possibility of erecting a complete consequentialist theory on kantian foundations remains, I believe, both appealing and important. But I won't attempt to sketch here what an adequate kantian theory of the good might look like. That must be left for another occasion.

Let me return, finally, to a point noted much earlier, when we first introduced FUL. Kant, it will be recalled, claims that FUL is itself only one way of stating the same basic imperative. That is, he held that there are other ways of formulating the very same categorical imperative in quite different language. For example, at one point in the *Groundwork* Kant claims that the categorical imperative can also be stated like this (the formula of humanity):

> FH: "Act so that you use humanity, as much in your own person as in the person of every other, always at the same time as end and never merely as means" (G 4:429).

And at a different point he claims that the categorical imperative can also be stated like this (the formula of the realm of ends):

> FRE: "That all maxims ought to harmonize from one's own legislation into a possible realm of ends as a realm of nature" (G 4:436).

It certainly must be admitted that it is far from obvious that these different formulas are truly equivalent, generating the very same guides to action. Indeed, not all kantians agree with Kant about the supposed equivalence. Of course, for that matter, it is also far from obvious how these alternative formulas are best understood, and how they are to be applied. Unfortunately, pursuing these related issues would involve considerable further discussion, and so we cannot consider them here.[18]

I do, however, want to address one final question that might naturally arise at this point. If there are different formulations of the categorical imperative, is there any particular justification for focusing, as I have, on FUL, as opposed to some of the alternative formulations? Perhaps not. After all, if they are genuinely equivalent, then they must all support the same moral principles. And if I am right in thinking that FUL may lead to consequentialism, then if they are equivalent the other formulas should lead to consequentialism as well. I find that a plausible claim as well, but I won't attempt to defend it here.[19]

But Kant himself, in surveying the alternative formulations of the categorical imperative, makes an interesting remark. With regard to some of the other formulas, Kant suggests, it might well be the case that they are more intuitive and accessible. But if we want a strict accounting of what to do, he says, then we should turn to FUL (G 4:436–37). I have followed Kant's lead in this regard, and focused on FUL itself. Of course, I have also argued that Kant may well be mistaken about where, precisely, FUL takes us. Kantianism, I have argued, represents a significant account of the foundations of ethics. But contrary to the claims of most kantians, and Kant himself, these foundations may well lead us to consequentialism.

NOTES

1. I mean here to distinguish between foundational theories and more "normative" theories — theories involving basic moral requirements such as those concerning harm doing, promise keeping, and so forth. For the distinction between these two levels of theory, see Shelly Kagan, "The Structure of Normative Ethics," *Philosophical Perspectives* 6 (1992): 223–242, or *Normative Ethics* (Boulder: Westview Press, 1998). I will have more to say about the distinction between deontological and consequentialist normative theories below.

2. Two important precedents for challenging this widely held view are David Cummiskey, *Kantian Consequentialism* (Oxford: Oxford University Press, 1996); and Richard Hare, "Could Kant Have Been a Utilitarian?" in his *Sorting Out Ethics* (Oxford: Oxford University Press, 1997). The former is particularly sensitive to the details of Kant's own position. But insofar as *Kantian Consequentialism* is primarily concerned with the formula of humanity rather than with the universal law formulation of the categorical imperative, the present essay is perhaps best viewed as being complementary to that work rather than simply duplicating it. I should perhaps note explicitly that while I have qualms about various details of Cummiskey's arguments, I am, of course, in broad agreement with his main conclusions. The same is largely true for Hare's discussion as well, though I am unconvinced that it is precisely *utilitarianism* — rather than some other consequentialist theory with a more complicated theory of the good — that emerges from Kant's account.

3. I will note, however, that Kant's own discussion of freedom is made complicated by his unargued assumption of incompatibilism — the claim that freedom is incompatible with determinism (see, e.g., G 4:446–47 or 455–56) — and that this is a view that the kantian need not accept.

4. Might the line of thought that leads from autonomy to FUL support an even stronger conclusion? If autonomy requires that I restrict myself to acting on reason-giving principles that I can autonomously will to be universal law, does it also require that I act on all those principles that I *can* so will? This is an important question, but I won't pursue it here (except to note that the distinction between what I can will, and what I do will, will be relevant). For simplicity, let's continue to follow Kant's lead and consider FUL only in its familiar, "negative" formulation.

5. His reasons for claiming this are not altogether clear or persuasive. At G 4:402 and 420–21 he seems to have in mind something like the following disjunctive argument: (1) the validity of imperatives must be based either on their content or on their form. But (2) considerations of content yield no categorical imperatives, and (3) the only categorical imperative based on form is FUL. So (4) the only categorical imperative is FUL. Now one worry about this argument is that it is difficult to see how to reconcile (2) with the later search (at G 4:428–29) for a formulation of the categorical imperative based on its inevitable content, a search that supposedly successfully results in the derivation of the formula of humanity. But since Kant holds that the formula of humanity is itself simply another way of formulating the same imperative as FUL, perhaps (2) could be replaced with (2′): the only categorical imperative derivable from considerations of content is equivalent to FUL. He could then still conclude with (4′) — that the only categorical imperative is FUL or its equivalent. The more serious difficulty with the argument, however, is that even if we grant (1) (and it is not clear that we should) neither (2) (or (2′)) nor (3) seems adequately defended or obviously correct.

6. Are there any such additional principles — valid, but not derived from

FUL? I don't see why the kantian should deny their existence. Indeed, as we will note later, many applications of FUL seem to make use of some sort of principle of instrumental reasoning. Kant defends his own favored version of this principle, but it is noteworthy that this defense doesn't make reference to FUL at all (see G 4:417). So there may be at least one such further principle, and I don't see why there shouldn't be others.

7. Kant speaks here of a universal law "of nature," since his discussion of the four examples actually proceeds in terms of the formula of the law of nature (FLN) — a variant of FUL which he introduces at G 4:421. For our purposes, however, the differences between FUL and FLN are unimportant.

8. See Christine Korsgaard, "Kant's Formula of Universal Law," reprinted in her *Creating the Kingdom of Ends* (New York: Cambridge University Press, 1996), for a fuller discussion of these and other interpretations, including a defense of the second.

9. Unfortunately, Kant seems to be confused on this point, sometimes apparently holding the view just shown to be mistaken — that the familiar moral rules are themselves *a priori* as well. (See, e.g., G 4:389, 408, or 410–12.) In any event, the claim that *FUL* is *a priori* is less clearly mistaken, and Kant certainly believed it too (see, e.g., G 4:419–21), though whether it is correct depends on, among other things, whether the autonomy of reason is something that can be established *a priori*.

10. This has an interesting implication, which I will mention in passing. People often take the familiar moral rules (to keep your promises, to tell the truth, and so forth) as themselves being categorical imperatives, binding upon everyone. But in light of what we have just noted, we must reject this view. (We would need to reject it in any event, if we insisted on taking seriously Kant's claim that FUL and its alternative formulations represent the *only* categorical imperative.) If the derivation of particular moral rules makes essential use of contingent empirical facts, then those rules will themselves only be binding *given* the facts in question. This means that moral rules will not be binding upon *all* rational beings, *regardless* of what else is true. Thus the familiar moral rules are not categorical — since categorical imperatives must be binding upon all rational beings without condition (see G 4:416). What *is* true, of course, is that they *are* binding, nonetheless, for those rational beings for whom the relevant empirical facts do obtain, and in a world like ours that may well mean for all human beings whatsoever. In particular, then, while the familiar moral rules are not categorical, they are not conditional upon the particular desires and goals of the people involved. (It must be admitted, however, that Kant himself seems confused on this point as well, suggesting at various places that the familiar moral rules are indeed categorical. See, e.g., G 4:389, 408, or 410–12.)

11. For one example of an argument to this effect, see Christine Korsgaard, "The Right to Lie: Kant on Dealing with Evil," in *Creating the Kingdom of Ends*.

12. Again, for the distinction between the two levels of moral theory see either "The Structure of Normative Ethics" or *Normative Ethics*.

13. I've argued elsewhere that, in general, deontological foundational theories need not support deontological normative theories (and, similarly, that consequentialist foundations needn't support consequentialist normative theories). See Part 2 of *Normative Ethics*.

14. Some deontologists are absolutists with regard to these constraints, holding the relevant types of acts to be forbidden no matter how much good would be done — or harm avoided — by performing them. Other deontologists are moderates about constraints, believing it permissible to infringe the constraint when enough good is at stake. For our purposes, however, the distinction won't be important.

15. Kant's own views on this subject are less clear. But it is striking that when discussing the aid example in terms of the formula of humanity Kant concludes that each person must "aspire, as much as he can, to further the ends of others" (G 4:430). This certainly looks like a denial of options. Perhaps, then, Kant would have agreed that FUL rejects options as well, since he believed that the formula of humanity is equivalent to FUL.

16. At least, not if we are taking the maxims to be *fully specified*. The consequentialist can readily admit, of course, that many other ("abbreviated" or "shorthand") maxims will also pass as well — when tested against the implicit background assumption that the act in question has good results overall.

17. Let me quickly mention another argument that is sometimes used to defend the claim that kantian foundations support deontology. It turns on the distinction between perfect and imperfect duties. A perfect duty, Kant says, "permits no exception to the advantage of inclination" (G 4:421 note) and is "unremitting" (G 4:424); in contrast, then, an imperfect duty presumably leaves one with some latitude as to how and when it is to be satisfied. Arguably this entails that one must never violate or otherwise infringe a perfect duty for the mere sake of fulfilling an imperfect one. If we then add the further assumption that the familiar duties such as the requirement not to lie, to keep one's promises, not to harm the innocent, and so forth, are perfect duties (because, supposedly, they all fall out of FUL at the

first step), while the duty to aid others is merely an imperfect duty (because it is generated only at the second step), we seem to have the desired deontological conclusion that one must promote the good, but not when this requires telling lies, harming the innocent, and so on. There are, however, a great many problems with this argument, not the least of which is the point, previously noted, that nothing in the account of FUL itself warrants investing the question of the stage at which a duty is generated by FUL with anything like this kind of significance. (For further discussion of the attempt to use the perfect/imperfect distinction as an argument for deontology, see Cummiskey, *Kantian Consequentialism*, chapter 6.)

18. Though I will note the obvious point that if one does accept more than one of these formulas as expressing a genuine categorical imperative, while denying their equivalence, one must deny Kant's claim, also previously noted, that there is exactly one categorical imperative. (Of course, it could still be the case that one of these categorical imperatives was the most basic, and the others could be derived from it.)

19. Though, again, see Cummiskey, *Kantian Consequentialism*, for a defense of the claim that it is actually consequentialism rather than deontology that is supported by the formula of humanity.