The Derivation of the Categorical Imperative: Kant’s Correction for a Fatal Flaw

Paul Guyer

In his chief works in practical philosophy, Kant derives the first and most famous formulation of the categorical imperative, the Formula of Universal Law expressed in such terms as “I ought never to act except in such a way that I could also will that my maxim should become a universal law,”¹ three times: in Section I of the 1785 *Groundwork for the Metaphysics of Morals*, from the common conception or rational cognition of duty; in Section II of the *Groundwork*, from the philosophical conception of a finite rational being as one who acts in accordance with a representation of a categorical imperative; and in the 1788 *Critique of Practical Reason*, from the conception of a fundamental principle of morality as a practical law. The gist of each derivation is the same: since any object or aim that might furnish the basis of a fundamental principle of morality is contingent, because it could be recommended only by contingent inclination, but such a principle must be necessary, the principle cannot concern the matter of action—that is, its intended object or aim—but only its form, that is, the universality that is inherent in the very idea of law itself; thus, all that is and can be required of any morally acceptable maxim is that it be universalizable. This argument fails for the simple reason that Kant does not consider the possibility that there might be a necessary object of action, recommended by something other than contingent inclination, that could serve as the basis of a universal and necessary principle of morality that is not purely formal in character. However, although Kant does not explicitly acknowledge this flaw in his derivation of the Formula of Universal Law, he nevertheless makes up for it when he introduces the idea of humanity as an end in itself, from which he derives the second formulation² of the categorical imperative, the Principle of Humanity: “So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.” (*Groundwork*, 4:429) Humanity as an end in itself is nothing less than a necessary object recommended by something other than contingent inclination, from which the necessity of acting only in accordance with universalizable maxims can in turn be derived. Kant’s stated reason for introducing the concept and Formula of Humanity is not to correct an acknowledged flaw in the preceding derivation of the Formula of Universal Law, but rather to begin progress towards his ultimate explanation of the possibility of our acting in accordance with the categorical imperative by reintroducing a conception of ourselves as both agents and patients of actions that could be our motivating ground for acting in accordance with the Formula of Universal Law. But it is a good thing that he does take this next step, for the fact that the Formula of Humanity entails the Formula of Universal Law makes up for the fact that the preceding derivation of the latter is a failure.³

In what follows, I will analyze each of Kant’s three derivations of the Formula of Universal Law (FUL) in order to demonstrate that they all overlook the possibi-
ty of a necessary object of action, and then show how the Formula of Humanity (FH) corrects this fatal flaw. I will not attempt here to provide an adequate treatment of the difficult question of just what does make humanity an end in itself.

I. The Derivation of FUL from the Common Concept of Duty

(Groundwork I)

In Section I of the Groundwork, Kant derives FUL from a conception of duty that he takes to be a matter of “common rational moral cognition” — this section is entitled “Transition from common rational to philosophic moral cognition.” (Groundwork, 4:393) Kant does not begin immediately with the concept of duty; he begins with his famous argument that a good will is the only thing that is unconditionally good. (4:393-6) But he then asserts that in creatures like us, who invariably suffer “under certain subjective limitations and hindrances” and do not automatically do what is necessary to achieve a good will, the task of achieving a good will presents itself as a duty. (4:397) Thus we can discover the principle upon which we must act in order to achieve a good will by discovering the principle upon which we must act in order to act from duty.

Kant proceeds to the statement of this principle by means of three propositions about duty that, in accordance with the title of the section, are intended to constitute the analysis of the common conception of duty, a conception we all acknowledge in our ordinary appraisals of actions (4:402) even if none of us ever actually lives up to it. The interpretation of this analysis is complicated by the fact that Kant never explicitly states the first of his propositions — we only know there is a first because he does explicitly state what he labels as the second and third propositions. (4:399-400) However, the content of the first proposition may be deduced from the examples by means of which Kant argues for it. What these examples imply is simply that an action from duty is never an action motivated by an inclination to perform it, where an inclination is any effect of feeling upon the faculty of desire.4 (see 4:412)

Kant reaches this conclusion by considering the role of inclination in four different relations of actions to duty. The first case, of course, is that of actions contrary to duty, which may be done from inclination (Kant does not explicitly say so), but in any case clearly are not done “from duty” (4:397) and therefore presumably will not by themselves reveal the principle of duty. There are then three different ways in which actions can be performed that are “in conformity” with duty, that is, instances of the type of action the performance of which duty requires in the relevant circumstances. First, an action in conformity with duty may be performed not out of any immediate inclination in favor of that action itself but on behalf of some further object to which the agent does have an inclination and which can be advanced by performing the action in question. An example of this is the case of a shopkeeper who does not overcharge his inexperienced customers, not because he has any immediate inclination against doing that, but because he has an inclination towards maximizing his profits, and he calculates that earning a reputation for fair dealing will be a means to realizing that end. Second, an action in conformity with duty can be performed out of an immediate inclination to do so, without any ulterior motive. An example of actions of this kind are acts of beneficence performed by
the "many souls so sympathetically attuned that, without any other motive of vanity or self-interest, they find an inner satisfaction in spreading joy around them." Finally, actions in conformity with duty can be performed for some reason other than any inclination, immediate or mediate, on their behalf, perhaps in the absence of any inclination whatsoever or even in the presence of contrary inclinations. An example of an action in conformity with duty performed in the absence of any inclination at all is a beneficent action performed by a philanthropist whose mind has been so "overclouded by his own grief" that it has "extinguished all sympathy with the fate of others." An example of an action in conformity with duty performed even in the presence of exclusively contrary inclination is the case of a person from whom "adversity and hopeless grief have quite taken away the taste for life," who "wishes for death and yet preserves his life without loving it, not from inclination or fear but from duty." (4:398) Kant’s claim is that only when actions in conformity with duty are performed from some principle of duty other than inclination, as they clearly are in the case of actions performed in the absence of any inclination or even in the presence of exclusively contrary inclination, do we ascribe such actions the special "esteem" or "moral worth" that we associate with actions performed from duty. "It is just then that the worth of character comes out, which is moral and incomparably the highest, namely that he is beneficent not from inclination but from duty." (4:398-9) The first proposition in the analysis of the common conception of duty must thus be that actions from duty are performed not from inclination, but from some other principle.5 Kant explicitly states the second proposition in his analysis of the common conception of duty:

An action done from duty has its moral worth not in the aim to be attained by it but in the maxim in accordance with which it is decided upon, and therefore does not depend upon the realization of the object of the action but merely upon the principle of volition in accordance with which the action is done without regard for any object of the faculty of desire. (4:399-400)

The first clause of this sentence might be taken to refer to the common practice of appraising actions for their intentions rather than actual outcomes, as when we praise people for trying to do the right thing even when they do not succeed or hold people responsible for intended murder even when they fail to injure their intended victims. But that practice by itself is compatible with virtually any normative moral theory, including consequentialism, because it can be understood as one of praising or blaming agents for the intended outcome rather than actual outcome of their actions in cases where we believe that factors beyond their control intervened between the former and the latter. Thus Kant’s conclusion that the moral worth of an action done from duty does not depend upon the realization of the object of the action but does depend upon a principle of volition that has no regard for any object of the faculty of desire at all could not be inferred from this common practice, for Kant’s assertion is equivalent to the claim that the moral worth of an action from duty has no connection even to the intended outcome of the action. Moreover, it is clear that Kant does not mean to infer his conclusion from this common feature
of moral appraisals, for he continues, “That the purposes we may have for our actions, and their effects as ends and incentives of the will, can give actions no unconditional and moral worth is clear from what has gone before,” (4:400) and “what has gone before” can only refer to the argument of the preceding paragraphs. But the argument of the preceding paragraphs is that the moral worth of actions done from duty is not dependent upon the presence of any inclinations. Thus, Kant’s argument must be that the moral worth of an action done from duty cannot depend upon the object, or the intended outcome, of the action because it cannot depend upon any inclination, an inference that holds only on the assumption that all objects of action depend upon inclination, that is, that an object of an action is always and only something that seems worth attempting to realize because of some inclination that it promises to gratify. Kant makes his assumption of this premise plain in the next paragraph when he states that “an action from duty is to put aside entirely the influence of inclination and with it every object of the will” (4:400): to put aside every object of the will is to put aside every inclination of the will only because all objects of the will are grounded in inclination. However, Kant fails to consider the possibility that there might be an object of the will suggested by something other than inclination, and thus fails to provide any argument for the assumption that is the key to his analysis thus far. If there could be an object of action suggested by something other than inclination, in particular a necessary object grounded in some necessary motivation rather than a merely contingent object grounded in contingent inclination, the principle of duty could turn upon an object for action without depending upon inclination.

Does Kant make this gap in his argument good in the third and final proposition of his analysis? No, he does not, for as he says, the third proposition, “duty is the necessity of an action from respect for law,” is intended to be simply “a consequence of the two preceding.” (4:400) I have in fact already quoted the premise for Kant’s third proposition and now just add his statement of his conclusion:

Now an action from duty is to put aside entirely the influence of inclination and with it every object of the will; hence there is left for the will nothing that could determine it except objectively the law and subjectively pure respect for this practical law, and so the maxim of complying with such a law even if it infringed upon all my inclinations. (4:400-1)

The third proposition adds no new premises to Kant’s argument, but simply draws the conclusion from the original proposition of the analysis and the intermediate inference from it: if the moral worth of an action from duty cannot be based on any inclination and therefore cannot be based on any object of action, there is no alternative but that it be based on some purely formal feature of the law in accordance with which the action is done and on the motivation of respect for that law. The entire argument depends upon the original assumption that inclination has no moral worth and the additional and undemonstrated assumption that all objects of action are grounded in inclination.

This is perfectly clear in Kant’s summation of the whole argument leading up to his first explicit statement of FUL in the *Groundwork*:
But what kind of law can that be, the representation of which must determine the will, even without regard for the effect expected from it, in order for the will to be called good absolutely and without limitation? Since I have deprived the will of all impulses that could arise for it from obeying some law, nothing is left but the conformity of actions as such with universal law, which alone is to serve the will as its principle, that is, *I ought never to act except in such a way that I could also will that my maxim should become a universal law.* (4:402)

Kant's argument is simply that since our conception of the moral worth of an action from duty implies that the performance of such an action must be independent of any and every impulse, immediate or mediate, that might give one a reason for conforming to any law; and then, since that in turn excludes "every effect expected" from the action, that is, every object of the action, from the grounds of its moral worth, there is no alternative but that only the form of the maxim of such an action, in particular its universality, be the source of its moral worth. But, again, this argument depends upon the assumption that all objects of actions are grounded in inclination, and Kant has failed, indeed he has not even attempted, to prove that this is true: he has made no effort to prove that there cannot be an object of action grounded in something other than inclination.

II. The Derivation of FUL from the Concept of a Practical Law

(*Critique of Practical Reason*)

At this point I turn to Kant's derivation of FUL in the *Critique of Practical Reason*, where FUL is in fact the only formulation of the categorical imperative that Kant introduces, because this argument so plainly makes the same crucial assumption as that of *Groundwork I*, although it starts from a different premise and uses more technical language. The argument of the "Analytic of Pure Practical Reason" in the second *Critique* is aimed at proving the "Fundamental Law of Pure Practical Reason:" "So act that the maxim of your will could always hold at the same time as a principle in a giving of universal law," in other words, the FUL (*Practical Reason*, §7, 5:30). The argument towards this conclusion begins with the "definition" or perhaps better, "declaration" (*Erklärung*), that "practical principles" are "objective, or practical laws, when the condition is cognized as objective, that is, as holding for the will of every rational being." (§1, 5:19) The argument then states that "All practical principles that presuppose an object (matter) of the faculty of desire as the determining ground of the will are, without exception, empirical and can furnish no practical laws." (§2, 5:21) What is empirical can furnish no practical law, of course, because the empirical is always contingent rather than necessary, and thus cannot hold for the will of every rational being: "a rule is objectively and universally valid only when it holds without the contingent, subjective conditions that distinguish one rational being from another." (§1, 5:20) But why is any principle that presupposes an object or matter for the will empirical or contingent? Because, Kant assumes, first, "the matter of the faculty of desire" is "an object whose reality is
desired;” second, “the relation to the subject...by which the faculty of desire is
determined...is called pleasure in the reality of an object;” and, third, “receptivity to
a pleasure or displeasure...can always be cognized only empirically and cannot be
valid in the same way for all rational beings.” (§2, 5:21) In other words, all objects
of the faculty of desire depend upon inclinations, or, as Kant puts it in the next sec-
tion, “All material practical rules put the determining ground of the will in the
lower faculty of desire,” (§3, 5:22) and inclinations cannot serve as the basis of a
practical law because they are contingent and do not hold for the will of every ratio-
nal being. The only alternative, then, is that “If a rational being is to think of its
maxims as practical universal laws, he can think of them only as principles that con-
tain the determining ground of the will not by their matter but only by their form.”
(§4, 5:27) In other words, it is only the fact that a maxim is fit to be a universal law
and not its object that allows it to satisfy the fundamental law of pure practical rea-
once again, the key assumption of Kant’s argument is that all objects or matter
of action are furnished by mere desires or inclinations, and, once again, Kant makes
no attempt to argue for this assumption itself by excluding the possibility that there
might be objects of desire suggested by something other than inclination, some-
thing that might be necessary rather than contingent and thus ground a necessary
object that could satisfy the condition for a practical law. Kant goes to some length
to argue that all pleasures or inclinations are pretty much the same, differing “only
in degree,” (§3, 5:23) and thus worth pretty much the same, so it is from the point
of view of pleasure indifferent and contingent whether one prefers reading or hunt-
ing, intellectual conversation or the gaming table. But he makes no attempt to
prove that all possible objects of the will are proposed only by pleasure or inclina-
tion, and thus his argument that the fundamental law of pure practical reason can
only be formal and not material fails.

III. The Derivation of FUL from the Concept of a Categorical Imperative
(Groundwork II)

The flaw in the derivation of FUL in the second section of the Groundwork
may be harder to see than the flaw in the argument of the second Critique, but it is
basically the same. While the first section of the Groundwork was supposed to be a
transition from common rational moral cognition to philosophical moral cognition,
which meant the derivation of a philosophically adequate formulation of the funda-
mental principle of morality from common conceptions of good will and duty, the
second section is supposed to be a “transition from popular moral philosophy to
metaphysics of morals,” (Groundwork, 4:406) which turns out to mean the replace-
ment of any popular moral philosophy with a derivation of the fundamental prin-
ciple of morality from a suitable philosophical concept. Popular moral philosophy is
the attempt to derive the principles of morality from empirical examples of human
conduct, and it leads either to the conclusion that there is no such thing as morality
at all, since it is not clear that anyone ever is motivated purely by duty, (4:407) or to
a hodgepodge of perfectionism, hedonism, divine command theory, and the like,
since people are in fact motivated by such things as interest in their own happiness
and fear of divine punishments. (4:410) The metaphysics of morals, by contrast, will
be the derivation of the fundamental principle of morality, or of its proper formula-
tion, from the analysis of "the universal concept of a rational being as such." (4:412)

This analysis proceeds as follows. First, Kant maintains that "Only a rational being has the capacity to act in accordance with the representation of laws," (4:412) a capacity that is therefore sufficient to identify rational beings although not, as it will turn out, to characterize them fully. A rational being is thus not simply one that acts in accordance with laws, which everything in nature does, but one whose own consciousness of the law upon which it is acting is a necessary condition of its action. "However, if reason solely by itself does not adequately determine the will" of a being with this capacity, that is, if the will of a rational being "is exposed also to subjective conditions (certain incentives) that are not always in accord with the objectives ones...then the determination of such a will in conformity with objective laws is necessitation" and the representation of an objective principle or law under this condition "is called an imperative." That "the will is not in itself completely in conformity with reason...is actually the case with human beings," (4:412-13) so in our case the representation of any law in accordance with which we have the capacity to act does take the form of an imperative. So just as before we could discover the principle of good will by analyzing duty, the form in which the good will presents itself to creatures like us who do not automatically have a good will, so now we can derive the principle from a consideration of imperative, the form that laws take for creatures like us whose wills, again, do not automatically act in accordance with reason.

Kant next claims that "all imperatives command either hypothetically or categorically." Hypothetical imperatives "represent the practical necessity of a possible action as a means to achieving something else that one wills (or that it is at least possible for one to will)." (4:414) In other words, they are reason’s representations of the means to desired ends, and they "necessitate" us or present themselves to us as constraints or imperatives for the simple reason that, although the principle that "Whoever wills the end also wills (insofar as reason has decisive influence on his actions) the indispensably necessary means to it that are within his power" (4:417) is analytic, our reason does not automatically have decisive influence on our actions, and sometimes we do not want to do what is necessary to achieve our ends. Kant goes on to distinguish two kinds of hypothetical imperatives, imperatives of skill that state the means to "problematic" or arbitrarily chosen ends, and imperatives of prudence that state the means to an end "that we can safely presuppose" that all human beings "actually do have by a natural necessity, namely happiness." (4:415) But since he then argues that happiness is not a determinate concept, but just consists in the satisfaction of whatever particular ends an individual happens to have chosen, which are all "problematic" or arbitrary, it turns out that there is actually no single imperative of prudence, (4:417-18) and the distinction between imperatives of skill and prudence can be ignored. The crucial contrast for Kant’s argument is that between the concept of hypothetical imperatives and that of a categorical imperative.

Hypothetical imperatives having been defined as those that represent the practical necessity of means to ends, Kant states that "The categorical imperative would be that which represented an action as objectively necessary of itself, without reference to another end." (4:414) Because such an imperative is without reference to
another end, Kant infers that “It has to do not with the matter of the action and what is to result from it, but with the form and the principle from which the action itself follows.” (4:416) This is the basis for Kant’s subsequent suggestion that “the mere concept of a categorical imperative may...also provide its formula,” and for his official derivation of FUL:

When I think of a categorical imperative I know at once what it contains. For, since the imperative contains, beyond the law, only the necessity that the maxim be in conformity with this law, while the law contains no condition to which it would be limited, nothing is left with which the maxim of action is to conform but the universality of a law as such...There is, therefore, only a single categorical imperative and it is this: act only in accordance with that maxim through which you can at the same time will that it become a universal law. (4:420-1)

Is there anything wrong with this argument?

One objection that leaps to mind is that by showing that there are two types of imperatives Kant has not shown that any rational being must act on both types of imperatives, and thus has not shown that any rational being must act on, or perhaps even be able to act on, the categorical imperative in particular. But it is clear that Kant does not intend to have shown that at this stage of the argument: that the moral law is actually binding on us and even “how such an absolute command is possible” is a “special and difficult” question that he explicitly states he will “post-pone to the last section” of the Groundwork (4:420). All he purports to be doing at this point is deriving the proper formulation of the fundamental law of morality as it presents itself to creatures like us.

But there is a problem in his execution even of this preliminary task, which is simply that he has not actually shown why an imperative that is categorical in any normal sense of the term must be “without reference to another end.” A standard interpretation of the characterization of a command as categorical would be that which is given by the first clause of Kant’s initial definition of the concept, that it is “objectively necessary of itself,” (4:414) or that which he employs in his introduction of the official derivation of FUL when he equates a categorical imperative simply with an “absolute command.” (4:420) The concept of a categorical imperative could also be equated with that of “an unconditional and objective and hence universally valid necessity,” (4:416) where if the term “unconditional” adds anything to the terms “objective and hence universally valid necessity” it would only be the thought that there is no exemption from an objective and universally valid necessity, no condition that can count as a ground or excuse for failure to satisfy it. To say that a command is categorical in this sense is simply to say that it is a law “that must be obeyed, that is, must be followed even against inclination.” (4:416) But how does it follow that a command that allows no exception must make no reference to any end or matter of action? It follows only on the assumption that no matter or end of action can give rise to a universal and necessary law, or conversely, that all ends or matters of action are merely contingent and subjective. And from what would this assumption in turn follow? Kant gives no argument for it here, so we can
only assume that his basis for it is what we have discovered elsewhere, namely the assumption that all ends of action are suggested by contingent inclination. But, as we have seen, Kant fails to exclude the possibility that there might be a necessary end of action mandated by something other than inclination. Kant’s equation of an imperative that commands absolutely and unconditionally with one that makes no reference to any end fails because he has not excluded the possibility of a necessary end that could also give rise to an absolute and unconditional command.

IV. The Alternative Derivation of FUL from FH

Kant follows the derivation of FUL and its illustration with four examples of duties that can be derived from it by observing that he has “not yet advanced so far as to prove a priori that there really is such an imperative” (Groundwork, 4:425); the question is still open, he says, whether “it is a necessary law for all rational beings always to judge their actions in accordance with such maxims as they themselves could will to serve as universal laws.” (4:426) The task seems to be to connect the necessity of the moral law to the will of a rational being: “If there is such a law, then it must already be connected (completely a priori) with the concept of the will of a rational being as such.” (4:426) But what is it about the concept of a will of a rational being as such that is not already included in the concept of a rational being as one that acts in accordance with the representation of a law that will allow this connection to be made? Kant’s assumption seems to be that it is a further constituent of the concept of a rational being, or of a rational agent, that is, a rational being with a will, that it always acts for the sake of an end, thus that it determines what law it will act in accordance with by reference to an end, thus that if it is to act in accordance with an objective law, it must act for the sake of an equally objective end, one that is valid for any and all rational beings: “Now, what serves the will as the objective ground of its self-determination is an end, and this, if it is given by reason alone, must hold equally for all rational beings.” (4:427) Far from acting with no reference to any end when it acts in accordance with the categorical imperative, then, a rational agent must act with a necessary end in view when it acts in accordance with this law and indeed determine itself to act in accordance with this law for the sake of this end: in such an end, “and in it alone, would lie the ground of a possible categorical imperative.” (4:428)

If by this point in the argument Kant had already established that FUL is the categorical imperative on some other ground, and if, as is ordinarily supposed, FUL is a test for the permissibility of all maxims proposing actions for the sake of any end,12 then the newly stated assumption that a rational agent must have some end in each of its actions would not imply that it must have a unique and necessary end in all of its actions, one that is valid for every other rational being. Any arbitrary end would give its action a point, or make it rational, and would be permissible as long as it passed the test of FUL. But clearly what Kant is supposing is that there must be a point to accepting FUL itself, something that makes it rational for any rational being to bind itself to FUL, and that this can only be a necessary end. He thus conceives of his present task as that of introducing such a necessary end, with the remaining task for the final section of the Groundwork then being that of showing how it is possible for us to adopt and act in accordance with this necessary end. The
introduction of such a necessary end could display the ground or reason for complying with a law the formulation of which has already been soundly derived. But if the first formulation of the law has not already been well established, as I have argued, then the introduction of the reason that makes it necessary to conform to this law should also be sufficient to derive the content of the law: it should be evident from the reason what it is a reason for. And this is exactly what now happens: Kant introduces as the necessary end of action for all rational beings an end that entails that all maxims must also be able to serve as universal laws, and thus in providing a reason for adherence to FUL makes good the failure of its previous derivation.

What is this necessary end? After stating that only in something “the existence of which in itself has an absolute worth, something which [is] an end in itself,” Kant simply proposes “Now I say that the human being and in general every rational being exists as an end in itself, not merely as a means to be used by this or that will at its discretion,” so every human being, “in all his actions, whether directed to himself or also to other rational beings, must always be regarded at the same time as an end” (4:428); this is what leads to the second formulation of the categorical imperative, “So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means.” (FH) (4:429)

As Kant subsequently says, “the formula of the moral imperative [which] is expressed thus: that maxims must be chosen as if they were to hold as universal laws of nature,” states only the “form” of moral maxims, while the formula “that a rational being, as an end by its nature and hence as an end in itself, must in every maxim serve as the limiting condition of all merely relative and arbitrary ends,” supplies the “matter” of moral maxims, that is, the end that is to be achieved through adherence to the form of moral maxims. Far from excluding all reference to any matter or end for actions, the formula of universal law excludes only contingent ends and depends upon the claim that rational being is a necessary end.

This naturally raises two questions. First, what is this “rational being” or its instantiation in us under the name of “humanity?” Second, why should it count as a necessary end, an end that can and must be “the ground of a possible categorical imperative” from which the content of such an imperative can indeed be derived?

Kant does not answer the first of these questions immediately. Several pages later, however, he states that “Rational nature is distinguished from the rest of nature by this, that it sets itself an end.” (Groundwork, 4:437) This implies that it is our capacity to set an end that is itself to be our ultimate end in all our actions. In the later Metaphysics of Morals (1797), he makes the same statement about “humanity” rather than “rational being:” “humanity (as distinguished from animality)” is the “capacity to set oneself an end—any end whatsoever” (“Doctrine of Virtue,” 6:392) or that “by which alone” a human being “is capable of setting himself ends” (6:387). On this definition, to make humanity always an end and never merely a means would make our capacity to set ends itself our necessary end even in our pursuit of more particular ends. If in characterizing humanity in this way, however, Kant were just appealing to some sort of empirical generalization about human beings, he would be undermining his pretension to derive the moral law not from some “special property of human nature” but from a concept that must “hold for all rational beings.” (4:425) But it seems clear that Kant does not intend to be offering a merely empirical generalization about human beings, and indeed he could be held
to be deriving this definition of humanity from his premise that a rational being must always have an end, or at least from a subtle transformation of this into the premise that a rational being is one that is capable of setting itself an end. Our humanity, as opposed to our animality, is the property that allows us, as opposed to the other animals, to be subsumed under the concept of rational being, so if our capacity to set ourselves ends is what makes us rational beings then that capacity is what constitutes our humanity.

Kant also makes it clear that it is our capacity to set an end that is itself our necessary end in his illustration of FH in the *Groundwork* by four examples of duty. Here he also makes it clear that what is entailed in the duty to make humanity itself an end and never merely a means is not just the preservation of the capacity to set ends, that is, the capacity for choice itself, but also the promotion of capacities to make this capacity effective when putting our choices into action. Although Kant’s discussion of the prohibition of suicide in the *Groundwork* does not make the point as clearly as it might, his discussions elsewhere make it clear that suicide is prohibited because, although it is the choice of an end, it destroys one’s own capacity for the further choice of ends: in suicide, freedom of choice “destroys and abolishes itself.” Making a false promise to another is wrong because, while not permanently destroying the other’s capacity to set his own ends, in the instance at hand it deprives him of the opportunity to exercise his capacity to set his own ends freely and in full cognizance of his actual circumstances: “he whom I want to use for my purposes by such a promise cannot possibly agree to my way of behaving towards him, and so himself contain the end of this action.” To treat the rational beings involved in any action “at the same time as ends” is to treat them “as beings who must also be able to contain in themselves the end of the very same action.” (Groundwork, 4:429-30) But further, making humanity an end in itself requires not just “the preservation of humanity as an end in itself” but also “the furtherance of this end,” (4:430) and this turns out to mean the promotion of conditions necessary to make the setting of ends by oneself and others effective. This can take the form of cultivating one’s own potential talents, precisely because “they are useful and are given for all sorts of possible aims,” (4:423) as well as that of directly assisting in the realization of particular ends chosen by others, which are worthy of realization precisely because they are “the ends of a subject who is an end in itself.” (4:430) Thus, rational being or humanity as an end necessary in itself is interpreted to require both the preservation of the capacity to set ends and the promotion of the effectiveness of this capacity in the realization of particular ends so chosen.

Suppose this is what FH requires. How does one get from this back to FUL, which I claim has not thus far successfully been derived? Kant clearly states that one can, when he says that

The principle, so act with reference to every rational being (yourself and others) that in your maxims it holds at the same time as an end in itself, is thus at bottom the same as the basic principle, act on a maxim that at the same time contains in itself its own universal validity for every rational being. (4:437-8)

And his treatment of false promises, for example, shows why this is true. If the
requirement to treat all the persons affected by any action as ends in themselves means that each such person must be able to make the action his or her own end, or set the end for that action by the exercise of his or her own humanity as the capacity to set ends, then every such person must be able to accept the maxim on which any agent proposes to act. But that means that any agent's maxims must be able to serve as maxims for all, or that FH implies FUL. In the first instance, to treat others as ends is to treat them as agents who must all be able to accept the maxims on which one proposes to act.\textsuperscript{16}

So FUL can be derived from FH, even if it has not successfully been derived before. Our second question, however, remains open, namely, why should we consider the capacity to set ends itself as our most fundamental end? This is perhaps the most vexing as well as most important question in the interpretation of Kant's moral philosophy. I cannot discuss it adequately here, and offer only some brief comments. One currently popular proposal is that the recognition that our capacity to set ends is supremely valuable is presupposed by our ascription of value to any particular end. Allen Wood, for instance, inspired by Christine Korsgaard, writes that "we have reason to regard as good the ends we ourselves set only to the extent that we (at least) implicitly respect and esteem our own rational nature as that which sets them."\textsuperscript{17} But it seems unlikely that this approach will work, for the simple reason that in order to infer the objective goodness of the capacity to set ends in each of us from the goodness of particular ends, each of us would have to recognize the particular ends of all of us as objectively good, and it is clear that Kant does not mean to start his moral theory from any such assumption. So what are the alternatives? One is that Kant simply supposes in 1785, as he did in 1764, that the "immediate supreme rule of all obligation must be indemonstrable,"\textsuperscript{18} thus that we will all simply acknowledge the force of this principle as soon as we reflect upon it. Another possibility, however, is that Kant meant to argue for this principle from some fact of common moral consciousness, just as he tried to argue for FUL from the common conception of duty. A candidate for such an argument would be that we can ultimately have respect only for the activity of an agent, and that it is only in setting our own ends, as opposed merely to having inclinations, that we are truly active, so that it is only the capacity for the activity of setting ends itself that is a proper object of respect. Kant hints at such an argument early in the \textit{Groundwork} when he writes:

\begin{quote}
Only what is connected with my will merely as ground and never as effect, what does not serve my inclination but outweighs it or at least excludes it altogether from calculations in making a choice—hence the mere law for itself—can be an object of respect and so a command. (4:400)
\end{quote}

Here Kant would assume not that inclinations are the only source of objects for action, as in his flawed derivations of FUL, but only that inclinations are passive, effects of other things upon our sensibility rather than acts of our will, and thus not suitable objects of respect. But our activity in setting ends, even when those ends might sometimes be suggested by inclination,\textsuperscript{19} is a proper candidate for respect.
V. Must the Right Precede the Good?

I will close by considering an obvious objection to my argument that although Kant’s interpretation of the categorical imperative as the Formula of Universal Law is not soundly derived prior to the introduction of the concept of humanity as a necessary end, it can be seen as not only motivated but also soundly derived from that conception. This is simply that Kant famously insists that the right must precede the good, or that any moral theory that infers what is right from an identification of the object of action will be heteronomous rather than autonomous. As Kant puts it at the end of the second section of the *Groundwork*:

> Wherever an object of the will has to be laid down as the basis for prescribing the rule that determines the will, there the rule is none other than heteronomy; the imperative is conditional, namely, *if* or *because* one wills this object, one ought to act in such or such a way; hence it can never command morally, that is, categorically. (4:444)

Why doesn’t this preclude the derivation of the categorical imperative from any conception of an object as a good, even from a conception of it as a necessary good?20

The answer to this question is just that Kant’s rejection of any derivation of the right from the good is based on nothing other than the assumption that all goods are suggested by inclination, which I have shown to be unjustified in Kant’s derivations of FUL and undercut in his argument for FH. That this is so may be somewhat obscured by the complicated explanation of the last statement that Kant offers in the *Groundwork*:

> Whether the object determines the will by means of inclination, as in the principle of one’s own happiness, or by means of reason directed to objects of our possible volition in general, as in the principle of perfection, the will never determines itself *immediately*, just by the representation of an action, but only by means of an incentive that the anticipated effect of the action has upon the will.... For, because the impulse that the representation of an object possible through our powers is to exert on the will of the subject in accordance with his natural constitution belongs to the nature of the subject—whether to his sensibility (inclination and taste) or to his understanding and reason, which by the special constitution of their nature employ themselves with delight upon an object—it would, strictly speaking, be nature that gives the law; and this, as a law of nature...is therefore in itself contingent and hence unfit for an apodictic practical rule, such as moral rules must be.... (4:444)

The initial contrast between inclination and reason might make it seem as if Kant means to consider and exclude objects of pure practical reason as well as objects of
sensibility from the grounds of any apodictic practical rule, and thus deny the pos-
sibility of any derivation of the right from the good. But it should be clear from
Kant's reference to the influence of an impulse on the will through understanding
and reason that he is referring to instrumental reason in the service of an inclina-
tion, and thus that he is saying only that inclination yields only a contingent end for
action whether it is the immediate source of action or only its mediate source. But
this simply fails to consider the possibility that there is a necessary end of action, set
by something other than mere inclination, which can be the source of an apodictic
practical rule.

The nature of Kant's assumption is even more evident in the Critique of
Practical Reason, where he writes:

If the concept of the good is not to be derived from an antecedent practical law but, instead, is to serve as its basis, it can only be the concept of something whose existence promises pleasure and thus determines the causality of the subject, that is, the faculty of desire, to produce it. Now because it is impossible to see a priori which representation will be accompanied with pleasure and which displeasure, it would be up to experience alone to make out what is immediately good or evil. (Critique of Practical Reason, 5:58)

This explicitly assumes that it is only subjective feelings of pleasure and displeasure, in a word, inclinations, which can furnish concepts of the good. But this is precisely what Kant failed to prove in his derivations of FUL and what is undercut by his identification of humanity itself as a necessary rather than merely contingent end. So Kant's thesis that the right cannot be derived from the good does not undercut the account of the role of the Formula of Humanity that has been offered here; it is only one more illustration of the assumption that Kant himself overcame in his argument for and from the Formula of Humanity.

Notes

1 With one exception, citations in this paper will be restricted to the Groundwork for the Metaphysics of Morals, the Critique of Practical Reason, and the Metaphysics of Morals, and will be located by their volume and page number as in Kant's gesammelte Schriften, edited by Royal Prussian (later German) Academy of Sciences (Berlin: Georg Reimer, later Walter de Gruyter, 1902- ). This citation is from Groundwork, 4:402. Translations are from Immanuel Kant, Practical Philosophy, ed. by Mary J. Gregor (Cambridge, U. K.: Cambridge University Press, 1996), modified where noted.

2 By designating the Formula of Humanity as the second formulation of the categorical imperative, I am declining to consider what is sometimes called the Formula of the Law of Nature, "act as if the maxim of your action were to become by your will a universal law of nature" (Groundwork, 4:421), as a separate formulation from the Formula of Universal Law. H.J. Paton famously argued that it is a separate formulation because it introduces a teleological conception of a law of nature that goes beyond its mere universality into the test of the morality of maxims; see his The Categorical Imperative (London: Hutchinson, 1947), ch. XV, pp. 146-57. Although Paton was correct that Kant's explanation of the prohibition of suicide in illustration of the Formula of Universal Law does
turn on a teleological conception of a law of nature that goes beyond its mere universality, he was wrong to suppose that Kant’s example was a proper illustration of his general principle. Kant explicitly states that the Formula of Universal Law can be reformulated as the Formula of the Law of Nature because “the universality of law in accordance with which effects take place constitutes what is properly called nature in the most general sense (as regards its form)” (4:421); in other words, to conceive of a maxim as universally acted upon just is to conceive of it as a law of nature, so testing whether one could will the universalization of one’s maxim is the same as testing whether one could will it to be a law of nature.

3 Allen Wood, in his recent work Kant’s Ethical Thought (Cambridge: Cambridge University Press, 1999), also argues that there is a “fallacy” in Kant’s deduction of FUL, namely, that Kant’s argument “tells us no more than that our maxims ought to conform to whatever universal laws there are. It does not tell us how to discover these laws...” (81). This is connected with his general charge that no particular positive duties, and perhaps even no adequately general negative duties can be derived from FUL alone, and thus that the derivation of duties from FUL presupposes an additional source of universal laws such as FH. My claim, however, is that there is a flaw in the derivation of FUL itself, not in the derivation of particular duties from it. In fact, I believe that Wood’s objections to the derivation of particular duties from FUL (v. 97-107) can all be rebutted, although that would be the topic for another paper.

4 In the introduction to their recent German edition of the Groundwork, Bernd Kraft and Dieter Schönecker assert that the first proposition in Kant’s analysis of duty is “An action from duty is an action from respect for the law;” v. Immanuel Kant, Grundlegung zur Metaphysik der Sitten, Philosophische Bibliothek 519 (Hamburg: Felix Meiner Verlag, 1999), xxi. But this goes beyond what can immediately be drawn from the examples by means of which Kant argues for the first principle, and instead equates the first proposition with the third, destroying the inferential structure of Kant’s analysis.

5 There is a vast literature on the question of how to interpret Kant’s examples. In the interpretation offered here, I am rejecting the assumption of Friedrich Schiller (see Wood, Kant’s Ethical Thought, 28) and many since that Kant’s view is that moral worth lies precisely in overcoming inclinations contrary to duty, or what Richard Hensen called the “battle-citation” model of moral worth (v. Richard Henson, “What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action,” Philosophical Review 88 [1979], 50), and instead accepting Barbara Herman’s position that for Kant “moral worth does not turn on the presence or absence of inclination supporting an action,” but rather on its exclusion from “the agent’s maxims as a determining ground of action;” see her “On the Value of Acting from the Motive of Duty,” originally in Philosophical Review 90 (1981): 359-82, reprinted in her The Practice of Moral Judgment (Cambridge: Harvard University Press, 1993), at p. 11. Having to overcome contrary inclination is closer to what Kant calls “virtue” than simply moral worth; v., e.g., Metaphysics of Morals, Introduction to the Doctrine of Virtue, 6:394 (Gregor, 524-5): “Virtue is the strength of a human being’s maxim in fulfilling his duty.—Strength of any kind can be recognized only by the obstacles it can overcome.”

6 Absicht. I have translated this as “aim” rather than “purpose,” as Gregor does, because the latter would usually be used as a translation for the different German word Zweck.

7 aller Antriebe. I translate this literally, in the plural, rather than in the singular, “every impulse,” as Gregor does.

8 In the Critique of Practical Reason, Kant does discuss the reformulation of the Formula of Universal Law as the Formula of the Law of Nature under the rubric “Of the Typic of Pure Practical Judgment” (5:67-71; Gregor, pp. 194-8). But he does not introduce the Formula of Humanity or any of the further formulations of the categorical imperative that are introduced in the Groundwork; the purpose of the second Critique is to amplify or revise the Groundwork’s explanation of how it is possible for us to act in accordance with the categorical imperative, that is, Kant’s theory of the freedom of the will, and not to amplify or revise the Groundwork’s formulation of the categorical imperative.

9 That is, the transition in the second section of the Groundwork is not supposed to be a transition from the conclusion of the first section to a further conclusion, but rather a transition from a different starting point to the same conclusion as that of the first section. The third section of the work is
then intended to effect a transition from the conclusion of both of the first two sections, the proper formulation of the categorical imperative, to a further conclusion, namely its necessary applicability to us as demonstrated by a critique of pure practical reason. The use of the term "critique of practical reason" for Kant's second Critique then confirms the claim that that the later work is meant to revise or amplify only the treatment of freedom of the will in the Groundwork, the subject of its third section, and not the whole work. For a similar analysis of the relation between the first two sections of the Groundwork, see Kraft and Schönecker, Grundlegung, xvi-xvii.

10 Various attempts have been made to assign meaning to this "through" (durch) which is not present in Kant's statement of FUL at 4:401. But since Kant's examples of the application of FUL (4:422-2) only require you to consider whether you could act on your maxim if it were to be universally acted upon, but do not suppose that your acting on your proposed maxim would itself cause it to be universally acted upon, it is not clear that this "through" actually means anything.

11 Kant's term is beurteilen. This is simply the transitive form of the verb urteilen, "to judge," which I prefer to Gregor's "appraise." See the editor's introduction in Immanuel Kant, Critique of the Power of Judgment ed. Paul Guyer, tr. Paul Guyer and Eric Matthews (Cambridge, U.K.: Cambridge University Press, 2000), xlvii-xlvi.

12 V., e.g., Wood, Kant's Ethical Thought, 100-7.


14 Translation modified.

15 In the Metaphysics of Morals, Kant argues that our imperfect duties, what he there calls ends that are also duties, can be only our own perfection, not our happiness, and the happiness of others, not their perfection ("Doctrine of Virtue," Introduction, 6:382-8). His argument for the first claim is that our own happiness is something we always naturally desire, not something we ever have to compel ourselves to pursue, and it is therefore not fit to be the object of a duty; his argument for the second claim is that the perfection of another consists precisely in his perfection of his own ability to set his ends in accordance with his own concepts of duty, and is therefore not something that can be the object of actions by anyone else. (6:386) Kant is wrong on both counts. As his own example of a sufferer from gout early in the Groundwork (4:399) implies, someone can easily put his short-term gratification ahead of his long-term happiness, and may actually have to pursue the latter out of duty rather than inclination; Kant also finally concedes that this is both possible and necessary later in the "Doctrine of Virtue" when he recognizes that "since all others with the exception of myself would not be all" (§27, 6:451), under appropriate circumstances the pursuit of my own happiness can be part of my duty to pursue universal happiness. Second, since the duty to preserve and promote humanity as an end in itself includes not just the duty to preserve the existence of the capacity to set ends and the possibility of its exercise, the latter of which indeed one cannot do for another, but also the duty to promote conditions that make the capacity of choice effective, one can help others develop the latter without undercutting their underlying freedom to set their own ends. Thus, one can have a duty to assist in the education of others without usurping their capacity to set their own ends.

16 For an account of universalizability as universal acceptability, see Nicholas Rescher, Kant and the Reach of Reason (Cambridge, U.K.: Cambridge University Press, 2000), ch. 8, 200-29.


19 I have argued that inclination must be recognized as the potential source of particular ends, even though inclinations can only be transformed into ends by the activity of reason, in several places, including "The Form and Matter of the Categorical Imperative" in Volker Gerhardt, Rolf-Peter Horstmann, and Ralph Schumacher, eds., Proceedings of the Tenth International Kant Congress
(forthcoming), and "Ends of Reason and Ends of Nature: The Place of Teleology in Kant's Ethics," *Journal of Value Inquiry* (forthcoming).