

The Supreme Principle of Morality

Allen W. Wood

1. What is 'The Supreme Principle of Morality'?

In the Preface to his best known work on moral philosophy, Kant states his purpose very clearly and succinctly: “The present groundwork is, however, nothing more than the search for and establishment of the supreme principle of morality, which already constitutes an enterprise whole in its aim and to be separated from every other moral investigation” (*Groundwork* 4:392). This paper will deal with the outcome of the first part of this task, namely, Kant’s attempt to formulate the supreme principle of morality, which is the intended outcome of the search. It will consider this formulation in light of Kant’s conception of the historical antecedents of his attempt.

Our first task, however, must be to say a little about the meaning of the term ‘supreme principle of morality’. For it is not nearly as evident to many as it was to Kant that there is such a thing at all. And it is extremely common for people, whatever position they may take on this issue, to misunderstand what a ‘supreme principle of morality’ is, what it is for, and what role it is supposed to play in moral theorizing and moral reasoning. Kant never directly presents any argument that there must be such a principle, but he does articulate several considerations that would seem to justify supposing that there is. Kant holds that moral questions are to be decided by reason. Reason, according to Kant, always seeks unity under principles, and ultimately, systematic unity under the fewest possible number of principles (*Pure Reason* A298-302/B355-359, A645-650/B673-678). Where systematicity is being given to empirical data, this may result in an irreducible plurality of principles, but the fact that that moral questions are to be

decided by reason gives us grounds for thinking that here there must ultimately be only a single principle. For this means that we must suppose there is an objective answer to them, an answer valid for all possible agents or inquirers (whether or not we are ever able to find that answer or agree on it) (*Groundwork* 4:442, *Morals* 6:207; cf. *Lectures on Ethics* 27: 276, 29:621, 625-626). It is familiar enough in everyday life, of course, that moral considerations are sometimes plural and mutually conflicting, but if there were no single principle to which they could be traced back, then necessarily there would be no objectively correct answer to moral questions whenever opposing answers could be made to rest each on its own ultimate, incommensurable principle. In that case is not even clear that we could consider the different answers conflicting answers to the same question, or consider there to be a specifically *moral* point of view, or even any determinate *moral* questions at all, since each of the irreducible plurality of principles would define a distinct practical viewpoint and a distinct set of practical questions, and no communication would be possible between these points of view concerning what, in the end, we *ought to do* or how we *ought to live*. This would spell the end of all moral objectivity, perhaps even of all morality, period.¹

That there is a supreme principle of morality, however, does not mean that there cannot be moral questions that are difficult to decide in practice, or that there must be an easy resolution to all moral conflicts and dilemmas. Nor does it mean that moral decisions are always, or even typically, to be made by referring them directly to the supreme principle. This is the mistake made by all those who think of Kantian ethics as recommending that we make all our decisions merely by applying Kant's famous formula

of universal law, asking ourselves “What if everybody did that?” Kant may have let himself in for such a mistaken reading when he said:

Thus I need no well-informed shrewdness to know what I have to do in order to make my volition morally good. Inexperienced in regard to the course of the world, incapable of being prepared for all occurrences that might eventuate in it, I ask myself only: Can you will also that your maxim should become a universal law?” (*Groundwork* 4:403).

However, the context of this remark must be carefully considered. Kant’s only aim in the passage is to draw a clear distinction between the prudential question whether it is safe to make a false promise for immediate gain and the moral question whether it is permissible to do so. He has just been observing that whether it is in our long-term self-interest to make a false promise is often a nice question, hard to decide on account of the conflicting considerations of momentary advantage and possible long-term risk. His point in this remark is that the same subtleties do not afflict the question whether it is morally right to make a false promise, since he thinks it is obvious that we could not rationally will that others should be allowed to perpetrate such deceptions on us, or fail to believe our promises -- as they obviously would if everyone were permitted to adopt the policy of making any promise they liked with no intention of keeping it. It is not at all clear, however, that the obvious generalization suggested by Kant’s remark is true, or is anything he would want to support. About many decisions made every day in the business world, for example (in particular, decisions about how far to be wholly frank with people and when to let them act on false beliefs) it is easy to see that these decisions are both safe and profitable, but a subtle and difficult question whether they are morally

right. We would seriously misunderstand Kant's ethics if we concluded from this passage that he has some deep theoretical reason for wanting to deny this obvious fact. The fact even further supports his main conclusion by showing another way moral questions can be easily distinguished from prudential questions.

Even more harmful and misleading, however, is the extremely common thought that Kant is recommending here that every decision we make in life should be prompted by asking ourselves whether some maxim or other can be willed as a universal law. This thought is responsible for so many misunderstandings, and there are so many things wrong with it, that it is hard even to know which ones to list first. This thought ignores the fact that, as we shall see below, the formula of universal law is only the first step in the process of formulating the supreme principle of morality, and consequently ignores Kant's other, richer and more definitive formulations of this principle. It does not consider that the formula of universal law provides only a negative test for maxims (a way of rejecting some as impermissible), but could never tell us in positive terms that we ought to follow any specific maxim. It disregards the fact that Kant never presents, and never uses, the formula of universal law as a general moral decision procedure. In any case, although the universalizability test may be suited to illustrate the specific examples to which Kant applies it, it would be radically defective as a general moral criterion, since it systematically yields both false positives and false negatives when we try to employ it generally.²

Against the general thought that the supreme principle of morality is to be used directly to make moral decisions, what is said by J. S. Mill might just as well have been said by Kant:

“It is a strange notion that the acknowledgement of a first principle is inconsistent with the admission of secondary ones...Men really ought to leave off talking nonsense on this subject, which they would neither talk nor listen to on other matters of practical concernment... Whatever we adopt as the fundamental principle of morality, we require subordinate principles to apply it by; the impossibility of doing without them, being common to all systems, can afford no argument against any one in particular.”³

In the case of Kant, he indicates clearly that the supreme principle of morality requires for its application a “practical anthropology” (*Groundwork* 4:388), so that we may determine what this highest principle -- so abstract and removed from ordinary decision-making that the search for and establishment of it must “constitute an enterprise whole in its aim and to be separated from every other moral investigation” – requires of us under the actual conditions of human life. When we turn to Kant’s actual account of ordinary moral reasoning in the *Doctrine of Virtue*, we see that it turns not on figuring out which maxims are universalizable, but on reasoning from a system of duties -- juridical and ethical, to ourselves and others, of respect and of love. Some of these duties are “perfect,” “narrow” or “strict”, requiring particular actions or omissions from us; most of them, however, are “imperfect,” “wide” and “meritorious”, requiring only that we set certain ends, and leaving it up to us to decide the priority among them and the specific actions that we will take toward them. Kant clearly recognizes that there can be conflicts between the different “obligating reasons” that arise from our various ends (MS 6: 224), and he worries a good deal (under the heading of ‘Casuistical Questions’) about cases in which

special circumstances might make it necessary to modify or make exceptions even to moral rules that are taken to be of strict obligation (see *Morals* 6:423-424, 426, 428, 431).

The role of a supreme principle of morality is not to dictate what we do in every particular case, but rather to stand behind and justify such a system of general moral rules or duties, and to provide a general rationale for deciding cases where reasons derived from them either collide, or leave it indeterminate what to do, or require us to make alterations in their demands to fit unusual situations. We would look in vain in the *Metaphysics of Morals* for any rigorous inferential route from the supreme principle of morality (in any of its formulations) to the specific duties Kant identifies. Only one of them is based on anything like the formula of universal law; all the others rest on appeals, usually both brief and casual, to the formula of humanity as end itself.⁴ But it is clear enough how the system reflects the general ideas of rational autonomy, the dignity of every rational being as an end in itself, and the laws by which every human being could rationally will that all should conduct themselves.

The function of a supreme principle of morality, then, is *not* to tell us directly, from day to day and minute to minute, through some uniform canonical process of moral reasoning to be applied in exactly the same way to all situations, exactly which actions we should (and should not) be performing and precisely how we should be spending our time. In this respect, we ought to ask far less of a supreme principle of morality than philosophers are in the habit of asking. But in another respect, we ought to ask a good deal more of such principles than is often asked. Analytical philosophers often aim at producing moral principles that may be very complex in structure, full of subclauses and qualifications, because these principles enable them to capture “our moral intuitions” and

the precisely worded epicyclic subclauses enable us to deal cleverly with threatened counterexamples of various kinds. (Kant's Formulas of Universal Law and the Law of Nature, when subjected to sophisticated interpretations that are intended to deal with all the troublesome counterexamples, are easily twisted into a principles of this kind.) But the resulting principles often do more to disguise than to state the fundamental value basis on which decisions are to be made. The right interpretation of Kant's formulation of the supreme principle of morality, by contrast, will be one that exhibits the principle as less concerned with generating results for all cases that accord precisely with our so-called 'intuitions', and more concerned with identifying perspicuously the ultimate value on which moral rules and duties may be grounded.

2. Formulating the Supreme Principle

In the *Groundwork*, Kant formulates the supreme principle of morality in conscious contrast to what he sees as the entire philosophical tradition of thinking on the topic. Further, in the twentieth century there was one interesting attempt to interpret the *Groundwork* as a conscious response to one influential historical text, namely Cicero's *On Duties*, especially as it had recently been interpreted by Kant's contemporary Christian Garve. But it will prove to be more perspicuous if we postpone such historical reflections until after an exposition of the procedure through which Kant develops his formulations of the supreme principle.

Duty and respect for law. Kant develops the moral principle twice in the *Groundwork*, using first a more commonsensical starting point in the First Section, then a more philosophical starting point in the Second Section, leading to a more complete

formulation. In the first section, the starting point is “common rational moral cognition.” The aim here is to enlist what Kant regards as certain of our most deeply held rational beliefs about morality on behalf of his new conception of the moral principle. He begins by focusing on the ‘good will’ which, he claims, we recognize as good in itself, and having a special place among goods in that it is the only thing good in itself whose goodness cannot be augmented or diminished by its combination with other good or bad things. Kant then attempts to forge a special connection between the good will and the idea of ‘acting from duty’ – that is, acting with inner rational moral constraint, motivated solely by the thought of following a moral principle. The crucial claim is that we think there is something uniquely worthy of esteem about a person who fulfills duty in the absence of (or even in opposition to) all other inducements of inclination or self-interest, solely out of respect for the moral law.

In light of over two hundred years of lively controversy over Kant’s assertions in the opening pages of the *Groundwork*, it is hard to resist the thought that Kant overestimated the extent to which the truth of his claims is available to all of us through ‘common rational moral cognition’. Our purpose here, however, is to see how he uses these claims to derive a formulation of the supreme moral principle. His central argument is that when we act from duty, even in opposition to all inclination, the only thing left that could motivate us is the purely rational appeal of a universally valid practical principle. This leads him to his first formulation of what we may call the Formula of Universal Law (FUL): “I ought never to conduct myself except so *that I could also will that my maxim become a universal law*” (*Groundwork* 4:402). In other words, the special motive of duty, which has a special affinity with the good will because it alone can rationally constrain us

to a course of action even in opposition to all our empirical desires or inclinations, can be nothing else but the unconditional worth of following a principle that binds us solely on account of its source in our own rational willing – in the fact that we regard it as a principle fit for being legislated to ourselves merely as rational beings, hence for being legislated universally to all rational beings.

Although Kant uses these thoughts only to reach FUL, they contain at least implicitly all the main ideas he goes on to develop, resulting in an entire system of different (yet, he argues, essentially equivalent) formulas of the supreme principle of morality. This more systematic exposition of the supreme principle of morality takes place in the Second Section of the *Groundwork*.

The Categorical Imperative. Crucial to the Second Section's formulation of the principle is the idea of a 'categorical imperative,' which can best be understood in connection with an entire philosophical theory of rational agency, presented very succinctly by Kant at *Groundwork* 4:412-421.

Kant's theory takes us to be agents who are self-directing in the sense that we have the capacity to step back from our natural desires, reflect on them, consider whether and how we should satisfy them, and to be moved by them only on the basis of such reflections. An inclination (or habitual desire we find in ourselves empirically) moves us to act only when we choose to set its object as an end for ourselves, and this choice then sets us the task of selecting or devising a means to that end. If I see an apple up in a tree and a desire to eat it occurs to me, then I will eat it only if I first decide to make eating it my *end*, and then devise a *means* (such as climbing the tree, or reaching for the apple with a stick, or knocking it to the ground by throwing something at it) to achieve the end.

In acting on my inclination, I thus make a series of decisions and create in myself a set of new desires (to climb the tree, or find a suitable stick) whose source is not merely the original desire I am trying to satisfy, but even more the exercise of my own capacities to set ends, devise means, and hold myself to some self-chosen plan for applying the means. Our desires, then, do not simply push us around like the levers and pulleys of a machine, but rather provide inputs into a rational process of self-direction involving our adoption and recognition of rational norms and the decision to follow or not follow the norms we recognize.

Setting an end is the most basic normative act, since (Kant holds) there is no action without an end to be produced by it. This act involves the concept of an object (or state of affairs) to be produced and also the concept of some means needed to produce it. Setting an end thus subjects me to a normative principle commanding me to perform the action required as a means to the end. Kant calls this principle a ‘hypothetical imperative’. It is called an ‘imperative’ because it is a command of reason requiring the agent to do something; it is ‘hypothetical’ because the command governs our action only on the condition that we will the end in question. By contrast, an imperative that has no such condition would be called a ‘categorical imperative’.⁵

Kant thinks that if the good will that acts from duty has the characteristic that it follows a rational principle even when all empirical incentives oppose it, then such a will should be understood as following a categorical imperative. For to act from duty is to follow a moral principle whether or not doing so achieves some antecedently desired end.⁶ Therefore, if acting from duty is what is most essential to morality, then the moral

law should also be characterized as a categorical imperative. Thus the supreme principle of morality, whatever else it is, must be conceived as a categorical imperative.

First Formula: Universal Law and the Law of Nature. As these considerations might lead us to expect, Kant now proceeds to derive essentially the same formula we saw at the end of the First Section, namely the Formula of Universal Law (FUL), which is now stated as: “*Act only in accordance with that maxim through which you can at the same time will that it become a universal law*” (*Groundwork* 4:421). By a ‘maxim’, Kant means a normative principle, which a subject lays down for itself with the intention of acting according to it. It perhaps involves a degree of idealization to represent agents as acting on ‘maxims’, since people do not typically recite to themselves (even silently) some general principle on which they are acting before they act. But the degree of idealization involved is not so great, when we consider that understanding an action at all normally involves understanding the agent’s intention, and the intention with which an agent acts is essentially such a subjectively adopted norm, usually also permitting us to form generalizations about what actions, consistent with this intention, the agent will perform or would perform under various counterfactual circumstances.

FUL provides us with a test for permissibility of maxims. It tells us that it is permissible to act only on those maxims we could will to be universal laws. The criterion of possibility here seems to be the absence of contradiction or conflicting volitions. It is not possible for me to will my maxim as a universal law if I cannot consistently think both of myself acting successfully on the maxim and also of its being a universal law, or if the volition that the maxim be a universal law would conflict either with the volition to

act on the maxim or else with some other volition that I, as a rational being, necessarily have.

The term ‘universal law’, as used in FUL, appears also to carry a normative force. That is, the question we are asking about our maxim is whether we could will that everyone (at least, everyone in our present circumstances) should be permitted to act on it. This is clearly the way Kant applies FUL in the First Section to the maxim of making the false promise: “Would I be able to say that anyone may make an untruthful promise when he finds himself in embarrassment which cannot get out of in any other way” (*Groundwork* 4:403). In other words, FUL invites us to consider which maxims we can will to be morally permissible for all, and commands us to restrict ourselves only to those maxims.

Apparently, however, Kant thinks it is easier (or more intuitive) to apply a different permissibility test to maxims, asking ourselves not which ones we can will to be universally *permissible*, but rather which ones we can will to be *actually followed as universal laws of nature*. (Again, the criteria of possible volition seem to be the absence of contradictions or conflicting volitions.) For in the Second Section, he immediately proposes this variant of FUL, which we may call the Formula of the Law of Nature (FLN): “*So act as if the maxim of your action were to become through your will a **universal law of nature***” (*Groundwork* 4:421). That is, we are to imagine a world in which, with the regularity of a natural law, the maxim we are considering is followed by everyone (in relevantly similar circumstances to the ones we are in). FLN, not FUL, is the formula Kant actually uses in illustrating his first formulation with reference to the four much discussed examples, organized according to the taxonomy of duties through which

Kant structures his more fully developed moral theory (that is, duties to ourselves and to others, perfect duties and imperfect duties).

Once Kant has completed his exposition of the supreme principle of morality, he tells us that the three formulas he has developed represent the moral principle from three different points of view: “form,” “matter,” and “complete determination” (*Groundwork* 4:436).⁷ The version of the first formula he identifies with “form” is again FLN: “That the maxims must be chosen as if they are supposed to be valid as universal laws of nature” (*Groundwork* 4:436). Both FUL and FLN may be regarded as identifying the “form” of the supreme principle of morality in the sense that they seek to specify a formal property of maxims such that having that property makes them compatible with the moral principle. This form consists in a certain relation to the rational will of the agent who proposes to act on the maxim, namely, the capacity of that agent to will that the maxim be a universal law of nature (or, in the FUL version, to will that its universal permissibility should be a valid norm for all rational beings).

Second Formula: Humanity as an End in Itself. Kant’s choice to begin by expounding the supreme principle of morality has been fateful regarding the misunderstandings and consequent (misguided) criticisms that it has provoked. Many, perhaps most, readers of the *Groundwork* have behaved as though Kant had intended his presentation of the moral principle to be complete at *Groundwork* 4:425. His further development of the supreme principle of morality in the *Groundwork* has been treated as a mere set of afterthoughts, not regarded as essential to interpreting the content of FUL and FLN or determining their role in Kant’s conception of the moral principle. It is no

exaggeration, however, to say that when the *Groundwork* is read in this way, its basic aims and contentions have been fundamentally misunderstood.

For example, Kant's entire approach to ethics has been (and still is) widely described as 'formalistic'. He has been criticized for not providing (or even for not allowing the possibility of) any substantive value lying behind the moral principle, or providing the rational will with any ground for being able to will one maxim, and not another, to be a universal law (or law of nature). The very concept of a categorical imperative has sometimes been rejected as nonsensical, on the ground that this concept precludes our having any substantive reason for obeying such an imperative. Schopenhauer, for instance, explained the alleged incoherence of Kant's thinking by attributing to him an ethics of divine command but without admitting a divine lawgiver to back up the command.⁸

Such criticisms are obviated, however, at least in the form they are usually presented, as soon as we turn from Kant's first to his second formulation of the moral principle. For it deals explicitly with the "matter" of the principle, by which Kant means the 'end' for the sake of which it is supposed to be rational to follow a categorical imperative. Kant's 'formalism' applies only to the first stage of his development of the principle; it is complemented immediately by considering the principle from the opposite "material" point of view, in which Kant inquires after our rational motive for obeying a categorical imperative, and locates this motive in the distinctive value that grounds morality, which he identifies with a kind of *end*.⁹

Here too, however, Kant's procedure was revolutionary, from a historical point of view, rejecting the standard picture of the kind of substantive value that might ground a

moral principle, and also the traditional conception of the sorts of things that can count as ends of human action. This radically new conception of the fundamental end of morality perhaps explains the incredulity that has often greeted the *Groundwork* on this point. The traditional view, namely, is that what grounds any principle must be an end to be produced, a state of affairs whose desirability gives us a reason to follow principles whose execution is conducive to bringing it about. As we have already noted, Kant accepts the traditional idea that every action has an end to be produced, but insists that the setting of such ends must be consequent on moral principles, not their ground. He rejects the thesis that any end to be produced grounds the supreme principle of morality, arguing that this would turn the principle into a merely hypothetical imperative, and deprive it of its status as a categorical imperative. The question then is: What sort of substantive value could give us reason to follow a principle without appealing to any end to be produced by following it?

Kant's answer to this question is found in the following remark, presented first in the form of a mere *supposition*: "But suppose there were something *whose existence in itself* had an absolute worth, something that, as an *end in itself*, could be a ground of determinate laws; then in it and only in it alone would lie the ground of a possible categorical imperative, i.e. of a practical law" (*Groundwork* 4:428). In other words, the substantive value grounding a categorical imperative cannot be the value of something future to be brought about as a consequence of our obeying it, but rather the value of something already in existence, which grounds our obedience to the imperative because such obedience serves to manifest or express our recognition of that value. Such an

existent value is an end in the sense that it is that *for the sake of which* it is rational for us to act.

Going beyond the mere supposition of something with this sort of value, Kant next presents his thesis in the form of an *assertion*: “Now I say that the human being, and in general every rational being, exists as an end in itself” (*Groundwork* 4:428). He then proceeds immediately to support the assertion by presenting, first, in a series of arguments eliminating other possible candidates for what might exist as an end in itself: the objects of empirical desires or inclinations, the inclinations themselves, nonrational beings (*Groundwork* 4:428). He follows this up with a brief, obscure but crucial positive argument that only ‘humanity’ understood in the technical Kantian sense of rational nature regarded as the capacity to set ends, can qualify as an end in itself. However we interpret this argument, the gist of it seems to be that we do rationally value our own existence as an end in itself, but we do so *rationally* only insofar as we value the existence of other rational beings in precisely the same way.¹⁰

Rational nature as an existent end in itself is distinct from all ends to be produced, but it stands in a determinate relation to them. All ends to be produced are set as ends by rational beings, since only rational nature has the capacity to regulate itself by rational norms, the most basic of which is the setting of ends and the selection of means to them (*Groundwork* 4:437). There are, in Kant’s theory, two basic kinds of ends to be produced that the supreme principle of morality requires us to set: our own perfection and the happiness of others (*Morals* 6:386-388, 391-394; cf. *Groundwork* 4:423, 430).¹¹

Regarded from the standpoint of its ‘matter’, then, the supreme principle of morality rests on the absolute worth of rational nature in the person of each human being,

and leads to the second main formula of the moral principle, the Formula of Humanity as End in Itself (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 an end, never merely as a means*” (*Groundwork* 4:429). As I have already mentioned, this is the formula of the moral law to which Kant most consistently appeals when he derives the duties belonging to the system he expounds in the *Metaphysics of Morals*.

Third Formula: Autonomy and the Realm of Ends. Kant has now derived two distinct formulas of the supreme principle of morality, both from the concept of a categorical imperative. The first was derived from the concept of a maxim that is compatible with this kind of imperative, and the general form that such a maxim would have to have. The second was derived from the concept of the substantive value (or the end) that could give us a rational ground to follow a categorical imperative. These two lines of argument from the concept of a categorical imperative are quite independent of each other, and lead to distinct formulations of the moral principle, even if (as Kant thinks) there is no conflict between these distinct formulas, and they can be treated as merely different ways of expressing “precisely the same law” (*Groundwork* 4:436). Kant’s next step, however, is to combine the two ideas behind these first two formulas to derive a third formula:

“The ground of all practical legislation, namely, lies *objectively in the rule* and the form of universality, which makes it capable of being a law (at least a law of nature) (in accordance with the first principle), but *subjectively* it lies in the *end*; but the subject of all ends is every rational being as an end in itself (in accordance with the second principle): from this now follows the third

practical principle of the will, as the supreme condition of its harmony with universal practical reason, the idea of the *will of every rational being as a will giving universal law*” (*Groundwork* 4:431).

The third formula combines the conception of a law valid universally for all rational beings (in FUL) with the conception of every rational nature as having absolute worth, to get the idea of the will of every rational being as the source of a universally valid legislation. The term ‘idea’ used in this formulation should be understood in Kant’s technical sense: an ‘idea’ is a concept of reason to which no empirical object can ever correspond, but which we use regulatively in arranging our cognitions in a system (*Pure Reason* A312-320/B368-377, A642-704/B670-732). Thus to regard the legislator of the moral law as the *idea* of the will of every rational being is not to say that the law is given by your arbitrary will or mine (for our wills are corrupt and fallible), but rather that the law is regarded as having been legislated by each of our wills insofar as it corresponds to an ideal rational concept of what it ought to be (but always falls short of being).

“The idea of the will of every rational being as a will giving universal law” is Kant’s initial presentation of the Formula of Autonomy (FA). It is also stated more directly, like the first two formulas, in the form of an imperative: “Do not choose otherwise than so that the maxims of one’s choice are at the same time comprehended with it in the same volition as universal law” (*Groundwork* 4:440). Or again: “Act in accordance with maxims that can at the same time have themselves as universal laws of nature for their object” (*Groundwork* 4:437). In these formulations, FA may sound superficially like FUL (or FLN), but in fact it is a formula quite distinct from either of

them, making a much stronger demand on maxims and yielding much stronger conclusions about what we ought to do.

Where FUL and FLN provide a mere condition of permissibility for maxims, consisting in its being *possible* (without contradiction or conflicting volitions) for you to will the maxim as a universal law, FA tells you *positively to follow* those maxims which *actually contain in themselves the volition* that they should be universal laws. FUL (or respectively, FLN) counts a maxim as permissible if there would be no contradiction or conflicting volitions in willing it to be a universal law (or law of nature); but a maxim might pass this purely negative test without containing in itself the volition that it should actually be a universal law (or law of nature). So the criterion on maxims proposed in FA is significantly stronger than the criteria of universalizability proposed in either FUL or FLN. And it justifies a correspondingly stronger conclusion about maxims, telling us not merely which ones are permissible and which not, but also which ones we have a positive duty to adopt because they are part of a system of universal moral legislation given by our own rational will.

Of course FA does not pretend to offer us any *test* to discriminate maxims that have this property from maxims that do not. But as I have already said, it would be error to think that the universalizability tests present in FUL or FLN are intended (even as permissibility tests) to apply to all conceivable maxims, so there is really nothing they can do that FA cannot. Both FUL and FA, rather, should be seen as indicating the spirit of a universal moral principle, and defining a task for reasoning: namely, in the case of FUL, that of deciding which maxims are compatible with a system of universal law (which maxims do not violate the laws of such a system), or, in the case of FA, which

ones belong to that system as part of its actual legislation as given by the idea of the will of every rational being.

FA combines in itself the main idea of FUL and the main idea of FA. Kant indicates this later when he says: “The three ways mentioned of representing the principle of morality are, however, fundamentally only so many formulas of precisely the same law, of which one of itself unites the other two (*deren die eine die anderen zwei von selbst in sich vereinigt*)” (*Groundwork* 4:436). This last clause has been mistranslated as “each of them unites the others in itself” or “any one of them of itself unites the other two in it.”¹² Both these translations say, as the original does not, that it is *equally* true of *each* of Kant’s three formulas that it unites the other two. However, it is only of FA that Kant ever explicitly claims that it unites the other two in itself; no such claim is ever made about FUL or FH. Consequently, I think we should regard FA as having a special status among the three formulas: FA is the formula that unites and sums up the others. It should be regarded as the definitive formulation of the principle of morality, insofar as there is one.

Just as Kant earlier provided a more “intuitive” version of FUL in the form of FLN, so here he also provides a more intuitive variant of FA, the Formula of the Realm of Ends (FRE): “Act in accordance with maxims of a universally legislative member for a merely possible realm of ends” (*Groundwork* 4:439). FRE provides a new characterization of the system of legislation referred to in FA, by describing the nature of the community that is to result from it. It calls this community a ‘realm of ends’ (*Reich der Zwecke*). By a ‘realm’ Kant means “a systematic combination of various rational beings through communal laws”, or again, “a whole of all ends in systematic connection” (*Groundwork*

4:433). In other words, a collection of ends constitutes a ‘realm’ if these ends are not in conflict or competition with one another, but are combined into a mutually supporting system. The laws of a realm of ends are those which, if followed, would bring the ends of rational beings (both the existent ends which are the rational beings themselves according to FH, and the ends set in the maxims chosen by those rational beings) into a mutually supporting harmony with each other. FRE commands us to follow maxims involving ends that belong to this mutually supporting system, and forbids us to adopt ends that fall outside it.

Kant sometimes looks upon this system (or “realm”) of ends as something like a single over-arching end, and thinks of following the principle of morality (as formulated in FRE) as joining with others in the shared pursuit of this collective end (or system of ends). The key terms Kant uses to express this idea are “system” (*System*) and “combination” (*Verbindung*). Thus at the conclusion of the *Anthropology*, he speaks of human progress from evil towards good as achievable only “through progressive organization of citizens of the earth in and to the species as one system, cosmopolitically combined” (*Anthropology* 7:333). Kant’s two main conceptions of what it is to act empirically according to the idea of a realm of ends are the relation of friendship, in which the happiness of both friends is “swallowed up” in a common end which includes the good of both, and the religious community, which in Kant’s view should be bound together fundamentally not by creeds or scriptural traditions but by the shared pursuit of the highest good as a common end.¹³

If this is right, then one interesting consequence is that FRE gives priority to securing human community or harmony over maximizing human welfare or satisfaction.

We should avoid all patterns of end-setting that involve fundamentally competitive relations between different rational beings, and we are forbidden to engage with others in ways that require the frustration of some people's deepest ends. Conflict or competition between human ends is permissible only if it is in service of a deeper systematic unity among all human ends, a system in which no member of the realm of ends is left out. The moral law commands us, in other words, to seek only that degree and kind of welfare for ourselves, and for others, that can be made to cohere with and support everyone's pursuit of the common welfare of all. If this means less total welfare than could be gotten by permitting fundamental conflicts between the ends of different rational beings, then lesser, not greater, total welfare is what the moral law commands us to seek.

The 'Universal Formula'. At this point, let us summarize the three (or five) formulas of the moral law the system of which constitutes the result of the *Groundwork's* search for the supreme principle of morality:

First formula:

FUL *The Formula of Universal Law: "Act only in accordance with that maxim through which you at the same time can will that it become a universal law"*
(*Groundwork* 4:421; cf. 4:402);

with its more "intuitive" variant,

FLN *The Formula of the Law of Nature: "So act, as if the maxim of your action were to become through your will a **universal law of nature**"*
(*Groundwork* 4:421; cf. 4:436).

Second formula:

FH *The Formula of Humanity as End in Itself: “So act that you use humanity, as much in your own person as in the person of every other, always at the same time as an end and never merely as a means” (Groundwork 4:429; cf. 4:436).*

Third formula:

FA *Formula of Autonomy: “... the idea of the will of every rational being as a will giving universal law” (Groundwork 4:431; cf. 4:432) or “Not to choose otherwise than so that the maxims of one’s choice are at the same time comprehended with it in the same volition as universal law” (Groundwork 4:440; cf. 4:432, 434, 438).*

with its more “intuitive” variant,

FRE *The Formula of the Realm of Ends: “Act in accordance with maxims of a universally legislative member for a merely possible realm of ends” (G 4:439; cf. G 4:433, 437, 438).*

As we have already noted, at *Groundwork* 4:436 Kant presents the three formulas as a system, characterizing FLN as giving us the “form”, FH the “matter” and FRE the “complete determination” of maxims under the moral law. He apparently chooses FLN over FUL and FRE over FA here because, as he says, his aim at this point is to “bring an idea of reason nearer to intuition (in accordance with a certain analogy) and through this, nearer to feeling” (*Groundwork* 4:436). He apparently means that FLN, using the analogy of practical laws with laws of nature, and FRE, characterizing the system of laws in FA through the analogy with an ideal community or realm of ends that is to result from it,

have greater appeal to us, thereby (as he elsewhere puts it) “providing entry” (into the human heart) for the precepts of morality (*Groundwork* 4:405). But after presenting his system of formulas with this intention, he points to the limits of his aim in the following remark: “But one does better in moral judging always to proceed in accordance with the strict method and take as ground the universal formula of the categorical imperative: *“Act in accordance with that maxim which can at the same time make itself into universal law”* (*Groundwork* 4:436-437).

The main point Kant seems to be making here is that the way of thinking (closer to “intuition” and “feeling”) that is best for animating human hearts and actions on behalf of morality is not the same as the way of thinking that is best when it comes time to pass critical judgment either on the actions we have performed or on the maxims we are proposing to adopt. For this latter task, apparently, a more austere and abstract principle is better, because, flawed human nature being what it is, the same feelings and intuitions that may make us enthusiastic friends of virtue also make us more susceptible to self-deception and make it easier for us to pass off corrupt actions and maxims as morally commendable ones. (In other words, those sentimentalists who think that what satisfies the heart, but not the head, represents greater moral purity, have things exactly wrong: where the head has been corrupted, it was the heart that corrupted it; and the first remedy for the corruption of our hearts is to learn to think in an enlightened way, with our heads, about what to do, and which feelings we should allow to influence us.)

In light of the systematization of the three formulations of the moral principle Kant has just presented, however, what are we to make of his reference to “the universal formula of the categorical imperative”? Is this intended to be the same as one of the other

formulas already derived? The most common interpretation is that the “universal formula” is FUL (perhaps because “universal formula” is, carelessly, thought to be (equally careless) shorthand for “formula of universal law”). Most who adopt this reading do so as if it were not the least bit problematic, as though it were simply what the text itself says.¹⁴ But of course it is not. I fear this is reflex reaction on the part of even many distinguished commentators is due to the pernicious influence of the traditional but deeply false idea that FUL is the primary (or in fact even the *only real*) Kantian formula of the moral principle.

Another (deeper and more interesting) thought, is presented by Klaus Reich (in an article whose main contentions we will be examining in the next section). This is that the “general” (or “universal” – *allgemein*) formula is yet a fourth (or a “sixth”) formula, distinct from all the “particular” formulas derived earlier in the Second Section and then systematized at *Groundwork* 4:436.¹⁵ This suggestion is interesting, and it gains some support from the fact that in Kant’s other two most important ethical works, *The Critique of Practical Reason* and the *Metaphysics of Morals*, the moral law is also represented by a single “universal formula” whose statements are very similar to that given at *Groundwork* 4:436-437: “So act that the maxim of your action could always at the same time hold as a principle of universal legislation” (*Practical Reason* 5: 30) and “Act upon a maxim that can also hold as a universal law” (*Morals* 6:225). But Reich’s suggestion raises the question where this new “general” formula is supposed to have come from, and in what way it is more “general” than the formulas already derived and explained.

Surely it is more natural to suppose, as the most common interpretation does, that the “universal formula” is one of the formulas already derived. The question, though, is:

Which one? There seem to me several reasons for thinking that it is to be identified not with FUL, but with FA. For one thing, the “universal” formula occurs in the same paragraph devoted to FRE (which is the more “intuitive” version of FA). Then too, as we have seen, FA is the formula that combines the other two in itself, and in which, in that sense, the search for the supreme principle of morality culminates. Further, the universal formula as presented in the *Critique of Practical Reason* is said reciprocally to imply freedom of the will (*Practical Reason* 5:28-30), but FA is the only formula in the *Groundwork* about which this claim is made (*Groundwork* 4:446-449). But the best reason is found simply in what the “universal formula” says: It tells us to act on that maxim that can *make itself* into a universal law. If a maxim “can make itself into a universal law” by “containing in itself the volition that it should be a universal law,” then this makes the “universal formula” equivalent to FA in several of its formulations. (If a maxim is able to “hold” or be “valid” (*gelten*) as a universal law whenever it contains in itself the rational volition that it should hold as a universal law, then the universal formulas of the moral law found in the *Critique of Practical Reason* and the *Metaphysics of Morals* can also be seen to be versions of FA.) By contrast, FUL tells us only to restrict ourselves to maxims that can (without contradiction or volitional conflict) be thought as universal laws; it does not tell us positively to act on maxims that can make themselves into such laws. The most compelling reading of *Groundwork* 4:436-437, then, is that the Second Section culminates in a system of mutually complementary formulas for the supreme principle of morality with each formula viewing the principle from a different standpoint. The universal formula, in which the others are combined and summed up, and which is the best standard to be used in moral judging, is FA.

3. Kant's *Groundwork* and Cicero's *De officiis*.

Having now examined Kant's attempt to develop a formulation of the supreme principle of morality, we turn next to a consideration of this attempt in relation to its historical antecedents. Our first task must be to evaluate the claim, which found considerable favor among some Kant scholars in the last century, that the *Groundwork's* formulation of the moral principle was consciously based on a particular ancient text, which was well known and influential in Kant's day, namely Cicero's treatise *On Duties*.

Kant probably began composing the *Groundwork* in 1783, after fifteen years of promising to write a 'metaphysics of morals'. In that year, the Berlin philosopher Christian Garve published a new translation of *De officiis*, and also a set of critical notes on it. Kant's brilliant but eccentric friend J. G. Hamann reports in correspondence that the philosopher began writing about moral philosophy about this time in order to provide an "anticritique" of Garve's book on Cicero, and then that by Spring, 1784 he was at work on a "*Prodromus der Moral*" (though terms like 'anticritique' and 'prodromus' sound more like Hamann's peculiar uses of language than they do like Kant) (*Akademie* 4:626-628). But these facts might lead us to wonder how far the *Groundwork* might have been influenced by Cicero's treatise *On Duties* (or by Garve's presentation of it).

During the twentieth century, reflections on this question led to some historical speculations about the genesis of some of the main ideas in the *Groundwork*, and the historical reference of Kant's formulations of the moral principle in it. Their source was an article by Klaus Reich, published in *Mind* in 1939. But Reich's speculations also influenced other scholars of the *Groundwork*, most notably H. J. Paton and A. R. C.

Duncan.¹⁶ Some of Reich's claims are quite plausible, such as that Kant was thinking of the classical list of virtues (justice, wisdom, courage and self-control), which he probably would have known about through Cicero, when he denies unqualified worth to both courage and self-control in the opening pages of the *Groundwork* (*Groundwork* 4:394; cf. Cicero, *On Duties* 1.15).¹⁷ Reich's most significant theses, however, concern the supposed sources in Cicero, and in the Stoic philosopher Panaetius of Rhodes, on whom Cicero was depending, for Kant's three main formulas of the moral principle as they are presented systematically at *Groundwork* 4:436. Specifically, Reich identifies FLN with the Stoic formula *convenienter naturae vivere* ("live according to nature") (Reich, p. 455; Cicero, *On Duties* 3.3), FH with Cicero's admonition that injuring another human being *omnino hominem ex homine tollit* (in Garve's translation, *im Menschen die Menschlichkeit aufhebt*, "abolishes humanity in the human being", Reich, p. 458; Cicero, *On Duties*, 3.5), and FRE with the Stoic formulas *communis humani generis societas* ("the society common to the human race", Reich, p. 459, Cicero, *On Duties* 3.5), *commune tanquam humanitas corpus* ("a community like a body of humanity", Cicero, *On Duties*, 3.6) and *deorum et hominum communitas et societas inter ipsos* ("the community and society of gods and men with one another", Cicero, *On Duties*, 1.43). Though Reich never quite puts it in this way, he writes as if in formulating the principle of morality, Kant was thinking of a series of Stoic formulations presented by Cicero early in Book Three of *On Duties* (and perhaps also of Garve's thoughts about them).

As Reich himself observes, there are no explicit references to either Cicero or Garve anywhere in the *Groundwork*. From this he rightly concludes that "in deciding on passages in which Kant took account of this work the greatest caution must naturally be

exercised” (Reich, p. 447). Duncan is if anything even more explicit: Reich’s conjecture, he says, is “no more than a hypothesis,” even a hypothesis that “cannot be established beyond reasonable doubt” (Duncan, pp. 178-179). The interest of this unprovable historical hypothesis, it seems to me, depends almost entirely on how much light it sheds on the philosophical content of the *Groundwork*. In other words: How much *philosophical* interest is there in the thoughts we entertain if we suppose that Kant’s formulations of the moral principle were inspired by reflections on the opening sections of Book Three of Cicero’s *On Duties*?

Judged by this criterion, I do not think Reich’s hypothesis fares well at all. There is (as Reich himself points out) a wide gulf separating the Stoic maxim “live according to nature” and FLN, which tells us instead to live according to laws we could will to be laws of nature. The thought that in injuring another I am removing or abolishing his humanity is not at all the same as, and it does little or nothing to illuminate, the thought that humanity, in the sense of rational nature, is an end in itself, and the fundamental value motivating obedience to all moral laws. (The comparative philosophical illumination of the two thoughts in relation to each other seems to be just the reverse: the Kantian thought would show why removing or abolishing someone’s humanity would be removing or abolishing something of great value. This might be implied by Cicero’s formulation, but it is not even explicit in it, much less subjected to philosophical elucidation.) There is certainly ethical as well as historical interest in the fact that the Stoics thought of humanity, or even gods and men together, as a single social body, but nowhere in this thought is there the crucial Kantian idea that the laws governing this body should be seen as proceeding from the idea of the will of each and every one of its

members, so that in obeying them, each is really obeying only himself. Regarding all three formulas, you need have the Kantian thought clearly in mind already before you can recognize anything like it in Cicero, and what you find in Cicero teaches you nothing at all philosophically about the essential Kantian thought.

We're no better off regarding the systematic connection between Kant's formulas. Although the quotations cited by Reich all occur within a relatively short space of text, as do the three formulas developed in the Second Section of the *Groundwork*, there is no suggestion in Cicero that these particular sayings constitute a single system defining the foundations of moral duty. On the contrary, they occur, along with many other thoughts, as part of Cicero's wide-ranging rhetorical argument in Book Three, whose main thesis is that there can never really be any conflict between rectitude or honor and mere expediency, but rather that the expedient or advantageous thing to do must always be the same as the right or dutiful thing to do (Cicero, *On Duties*, 3.4).¹⁸ On this point, however, far from its being true that Kant might have been inspired by Cicero, it would be no exaggeration to say that the emphatic repudiation of Cicero's thesis is one of the most persistent themes throughout the entire *Groundwork*. Yet it is hard to convince oneself even that Kant was setting out to argue against Cicero in particular here. For in the *Groundwork* there are no references to Cicero's defense of this thesis, and no discernible attempt to address any of his particular arguments in favor of it.

Thus looking at the matter from every point of view, and even supposing for the sake of argument Reich is correct in conjecturing (on the basis of no real evidence worthy of the name) that Kant had Cicero's treatise in mind while he was composing the *Groundwork*, it still seems that the argument of the *Groundwork*, regarding what is

philosophically interesting in it, proceeds very much as if Kant had not been thinking about Cicero or Garve at all. In other words, we learn virtually nothing of philosophical interest about Kant's formulation of the supreme principle of morality from reflecting on this fact about his private mental history (again assuming, with no explicit evidence, that it is a fact). This makes it very hard to concur with Duncan's insistence that "No one who undertakes to write about Section II [of the *Groundwork*] can afford to neglect [Reich's article in *Mind*]" (Duncan, p. 175). Or at least, speaking only for myself, I must confess that I have learned practically nothing about the philosophical content of the *Groundwork* by attempting to reflect on Reich's unprovable historical speculations. (If others more discerning or imaginative than I am are capable of finding these conjectures more philosophically illuminating than I have, then more power to them.)

If we are to understand the relation of Kant's search for the supreme principle of morality to its historical antecedents, I think we would do better to look at this in light of Kant's own explicit statements about it – in the *Groundwork*, in the *Critique of Practical Reason*, and elsewhere. That is what we will do in the final section.

4. Kant's Critique of Previous Attempts to Discover the Supreme Principle of Morality

Ancient ethics did not talk much of 'moral principles', but oriented itself either toward conceptions of human virtue or conceptions of the human good. During the twentieth century, there arose among moral philosophers a fashionable view – combining, with an appearance of inconsistency that merely added to its attractiveness, the appeal of iconoclasm with that of piety toward what is old – that this point represents an advantage of ancient over modern ethics, perhaps even showing that there is

something misguided about modern moral philosophy as a whole. According to one seminal and influential presentation of this view, that of G. E. M. Anscombe, the modern conceptions of ‘obligation’, ‘ought’ and ‘moral principles’ are dependent on a “law conception of ethics,” which in turn makes sense only within a religious view of the world in which God is thought of as the moral lawgiver.¹⁹ But it is precisely this conception, she claims, that modern moral philosophy has given up. So it finds itself working with a set of conceptions that, apparently through some sad fit of absent-mindedness on the part of virtually all modern moral philosophers, have managed to survive “the framework of thought that made [them] really intelligible” (Anscombe, p. 31).

Since Kant’s moral philosophy would seem to be sitting right in the bull’s-eye of the target at which such polemics are aimed, it might surprise their proponents to learn that he accepts the historical side of their contentions, at least up to a point.²⁰ According to Kant, the question of “the basis of morality” which asks about “the principle of morality,” “has been investigated in the modern age” (*Lectures on Ethics* 29:620). In place of this, by contrast, he says, the ancients asked about the *summum bonum*, the highest good (*Lectures on Ethics* 27: 247, 29:599; cf. *Practical Reason* 5:111ff.).

Ancient ethics: “The ideal”. In Kant’s view, the highest good was conceived by the ancient schools in a variety of ways. All of them were oriented primarily to “the ideal,” that is, the “pattern, idea or archetype” of what a human being can be. For some (but not all) of the ancients, the ideal was also associated with a conception of happiness. Among the ancients, Kant distinguishes the following theories of the ideal:

1. The Cynic ideal (of Diogenes and Antisthenes), which is *natural simplicity*, and happiness as the product of nature rather than of art.
2. The Epicurean ideal, which is that of the *man of the world*, and happiness as a product of art, not of nature.
3. The Stoic ideal (of Zeno), which is that of the *sage*, and happiness as identical with moral perfection or virtue.
4. The mystical ideal (of Plato), of the visionary character, in which the highest good consists in the human being seeing himself in communion with the highest being.
5. The Christian ideal of holiness, whose pattern is Jesus Christ. (*Lectures on Ethics* 27:247-250; 29:602-604).

The first three ideals place the incentive to morality in happiness, but the last two do not (*Lectures on Ethics* 27:250). The Cynics and Epicureans think of happiness as an *effect* of achieving the ideal (hence they think of moral virtue -- conceived alternatively as natural simplicity and worldly wisdom -- as a *means* to happiness), while the Stoics think that happiness is *identical* with achieving the ideal (*Lectures on Ethics* 27:250-251; *Practical Reason* 5:111-112).

There is another general criticism that Kant addresses against ancient ethics as a whole: that it fails to distinguish principles of right from those of morality, and treats both under the common heading of ethics. It is noteworthy that Cicero's *On Duties* that Kant selects as the chief target of this criticism (*Lectures on Ethics* 27:481-482). Perhaps it is even more noteworthy that the position Kant is here criticizing in the ancients is one frequently attributed nowadays to Kant himself, by those who hold that for Kant right is

subordinated to ethics and that Kant's principle of right is derived from his principle of morality.²¹

Modern ethics: Principles of morality. The highest good may have seemed to the ancients like a natural starting point for ethics, but Kant thinks that it treats as primary what is really secondary, and more fundamentally, it fails to ask the basic question. For, he contends, "the concept of good and evil must not be determined before the moral law (for which, it would seem, this concept would have to be the foundation), but only... after it and by means of it" (*Practical Reason* 5:63). The moderns, therefore, are asking the right question in inquiring after the basis of morality in its supreme principle. Some of their answers to this question, in his view, are non-starters because they substitute mere analytic judgments for a principle that must be synthetic if it is to ground the activity of practical reason. Other answers are faulty because they have not separated themselves far enough from the ancient standpoint. And all previous answers remain unsatisfactory to the extent that they have proposed a basis for morality in principles of heteronomy.

Analytic principles. Kant considers several ethical principles that he rejects because they attempt to pass off an analytic judgment as if it were more than that:

1. Do good and avoid evil. (Wolff).
2. Act according to the truth (Cumberland).
3. Act according to the mean between vices (Aristotle). (*Lectures on Ethics* 27:264, 276-277).

'Do good and avoid evil' is trivial because the concept of a good action is simply that of an action that is to be done, and the concept of an evil action is that of one that is to be omitted. The principle attributed here to Richard Cumberland is actually one that is held,

in various forms, by virtually all adherents of the British rationalist tradition in ethics, including Ralph Cudworth, Samuel Clarke, William Wollaston and Richard Price.²² It holds that actions have a real nature, and are involved with real relations to things and to other actions. In virtue of these natures and relations, it is *true* of some actions that they are *right* or *to be done*, and of others that they are *wrong* and *to be avoided*. Presumably Kant's criticism of the principle that one should act in accordance with such *truths* is that this principle actually says no more than Wolff's principle does (for it tells us only to perform those actions of which it is *true* that they are right and ought to be performed). It is curious that Kant should have listed Aristotle's principle of the mean along with principles of the moderns, and curious also that Aristotle finds no place in Kant's account of the ancient schools. But his criticism is no doubt that, like Wolff's principle, it tells us only to do those actions that fall under the concept 'to be done'.²³

Principles of Heteronomy: Kant's taxonomy. Kant criticizes the moral principles proposed by previous philosophers by characterizing them as 'heteronomous' in contrast to his own principle of autonomy (in the form of FA). In the *Groundwork*, his taxonomy of such principles distinguishes, "rational" principles from "empirical" principles; in the *Critique of Practical Reason*, the same distinctions are made, but this time between "determining grounds" of moral principles, and the distinction is between "objective" (instead of "rational") and "subjective" (instead of "empirical") grounds. There, each of these groupings is further divided, in a way that cuts across this first distinction, into "external" and "internal" grounds. The second *Critique's* taxonomy (*Practical Reason* 5:40) thus looks like this:

Subjective

External
Education
(Montaigne, [Mandeville])

Internal
Physical feeling
(Epicurus, [Hélvetius])

Civil Constitution (Mandeville, [Hobbes])

Moral feeling
(Hutcheson, [Shaftesbury])

Objective

Internal
Perfection
(Wolff, the Stoics, [Baumgarten, Cumberland])

External
The will of God
(Crusius, the theological moralists,
[Baumgarten])

(The names inserted in brackets are found in Kant's lecture presentations of the distinctions, not in the *Critique of Practical Reason*; see *Lectures on Ethics* 27:253, 510, 29:621-622, 625-627.)²⁴

It is significant (and clarifying) that in the *Critique of Practical Reason*, Kant treats this as a taxonomy not of moral *principles* but of moral “determining grounds” – that is, of proposed *grounds for following* moral principles. The association between the two is in any case quite natural in most cases. To view (with Montaigne) custom and education as providing the source of moral duties naturally goes along (at least in Kant's view) with viewing the moral ground as “imitation” or “example” – morality consists in doing what one has been taught by custom to do simply because it is what others in one's society do (and are approved for doing); likewise, to think (with Mandeville or Hobbes) of moral principles as those legislated in civil constitutions is to treat the coercive force of the sovereign as one's ground for following them (*Lectures on Ethics* 27:253, 29:621). To make the moral principle happiness or moral feeling is to treat the desire for happiness or moral sentiments of approval and disapproval as the ground for complying with these principles. Theological ethics, making God's will the moral principle, treats either fear or

love of God as the ground of morality, while the principle of perfection takes the value of perfection as our ground for following the principle.

Kant's critique of principles of heteronomy. The shift is also significant because Kant's critique of these alternative principles, both in the *Groundwork* and elsewhere, also focuses essentially on issues raised by these grounds for following them. Crucial to Kant's criticism of these alternative principles is his claim that each of them takes some object of the will as the determining ground of the rule that is to govern the will. The principle of education takes the imitation of examples as such an object, the principle of civil constitution, the fear of the sovereign's sanctions, the principle of physical feeling takes the enjoyment of pleasure, the principle of moral feeling takes the feeling of approbation (or the avoidance of feelings of disapprobation), the theological principle takes conformity to God's will, the principle of perfection takes the achievement of perfection. But, Kant argues, if the determining ground is an object of the will, then the imperative grounded on it can be only hypothetical, since the validity of the imperative for the will is conditional on achieving that object:

“Wherever an object of the will is to be taken as the ground in order to prescribe the rule determining that will, there the rule is nothing but heteronomy; the imperative is conditioned, namely: *if* or *because* one wills this object, one ought to act thus or so; hence it can never command morally, i.e. categorically” (*Groundwork* 4:444).

This argument needs to be understood against the background of what Kant thinks he has already established in the *Groundwork*, namely, that if morality is not a mere cobweb of the brain, then its supreme principle is a categorical imperative, and such a principle can

be comprehended, in its most developed and universal form, as autonomy of the will (FA). The ultimate value on which this principle rests is the dignity of the rational will as capable of giving universal law to itself and to all other rational wills. The advantage of the principle of autonomy is that it enables us to conceive the validity of the moral principle as independent of any object of the will. All objects of the will (such as the “ideals” of ancient ethics, or their conceptions of happiness, or any of the objects providing the determining grounds involved in the taxonomy of modern ethical principles) are thereby shown to be inadequate grounds for morality, in contrast to the principle of autonomy, which alone can be made consistent with the idea of a categorical imperative.

Once the force of this argument is appreciated, it is easy to understand why partisans of the various ethical principles Kant rejects should react to it by attempting to discredit the very concept of a categorical imperative. For in light of this argument, that concept seems to set up a hurdle that their favorite principle can never jump. The most obvious first reaction is therefore to criticize the demand itself as unreasonable and the concept supporting it as nonsensical. However, contrary to this first reaction, we can see fairly easily that none of these theories turns for its defensibility on the question whether the notion of a categorical imperative makes sense. For, as I will now argue, those principles that *must* hold that it does not are indefensible even if we reject that notion, while the rest can, contrary to Kant’s contention, meet his demand that they be understood as categorical imperatives. Empiricist theories, namely, are hopeless even apart from Kant’s criticisms, while rationalist theories are quite defensible against them.

The strategy of denying that the notion of a categorical imperative makes sense is the only one available to those defending *empirical* principles, or *subjective* determining grounds for the moral principle. For they are committed to saying that in the end, the only reason we can give for following the moral principle is that we are *so built by nature* that we have certain *desires* (to imitate others, to avoid the sanctions that the sovereign might impose, to feel pleasure). Or at most, they can say that we are so built that we *count* something *as a reason* (a feeling of approval or disapproval). Such empirical desires or dispositions to take something as a reason are necessarily only contingent features of our nature, without which the principle in question would have no validity for us. At most, then, they could supply us only with hypothetical imperatives. Partisans of such views often announce this point themselves, insisting that to ask for more than this is to indulge in metaphysical nonsense. (You can always tell when the hollowness of an empiricist view is in danger of exposure by the empiricist's desperate resort to the accusation that you are committing 'metaphysics'.)

Yet the problem with such views seems to me to go even deeper than Kant's criticism reveals. For even if we do not insist that moral principles are categorical imperatives, we ought at least to insist that there must be *some genuine reason* (categorical or not) for us to follow them, and none of the empirical theories seem consistent with meeting even that minimal requirement. For the fact that we are so built that we desire something does not give us a reason to desire it, nor a reason to satisfy the desire that we may have for it. Nor does the fact that we are so built that we take some feeling to be a reason for doing something amount to there really being a reason for us to do it. For as rational beings, we are also so built that we are capable of requiring genuine

reasons for doing what we do, and also capable of recognizing bogus substitutes for reasons as bogus.²⁵ (If it takes ‘metaphysics’ to acknowledge that there are genuine reasons, then that is about as good a defense as ‘metaphysics’ could ever hope for, since then it would then be self-contradictory for anyone to claim they had a genuine reason to reject ‘metaphysics’.) The empirical theories are therefore indefensible even apart from the Kantian worry that they cannot treat moral principles as categorical imperatives.

The rational principles are not so badly off. In fact, I do not think that rationalists necessarily need to attack the idea of a categorical imperative in order to save themselves from Kant’s criticism. Theological moralists (at least a certain kind of rationalistically minded theological moralist) may say that we are obligated to obey the divine will because that will is perfect, and hence what it wills or commands really is right in itself, independently of whether or not obeying the command achieves any further object of the will. But, if we allow the notion of a categorical imperative, then that is just to say that the commands of a perfect (divine) will are categorical imperatives. Likewise, a defender of the principle of perfection may say that this principle means only that we have a reason to act according to the idea of a perfect will, simply because perfection of will is intrinsically good -- again quite apart from whether so acting achieves any other object. That allows the perfectionists to say that their principles are categorical imperatives (again, if assuming we accept the idea of a categorical imperative). The British rationalist variant of this is that we are obligated to do those actions whose nature marks them out as right or to be done, while we are obligated to refrain from those actions whose nature makes them wrong or not to be done.²⁶ Again, the reason for doing and refraining lies in the nature of the actions themselves, and is not dependent on whether the doing or

refraining achieves any other object of the will. There is nothing in this that is inconsistent with regarding the principle of morality as a categorical imperative.²⁷ Ancient “ideal” theories of ethics may be defended in the same way, as long as their ideals are interpreted in a rationalist rather than an empiricist way. This is perhaps unpromising for the Epicurean ideal (and probably also for the Cynic ideal) since they seek either worldly virtue (or natural simplicity) as means to happiness, but can in principle give no account of why we have a genuine reason to want to be happy. But it seems quite possible for the Stoic, mystical (or Platonic) and Christian ideals to be framed in terms that are compatible with understanding morality as grounded on a categorical imperative. With the ancients, then as with the moderns: rationalism is defensible, empiricism, indefensible, whether or not we decide the notion of a categorical imperative makes sense.

Thus we arrive at the following conclusion concerning Kant’s critique of the conceptions of the supreme principle of morality that preceded his own in the history of ethics: Kant is right in rejecting empirical moral principles, but he does not need to assume the idea of a categorical imperative in order to do so, for they are quite hopeless even apart from that idea; on the other hand, his arguments do not necessarily discredit rational moral principles, since these can be so understood that they can be just as easily brought into harmony with the idea of a categorical imperative as can Kant’s own principle of autonomy.

Notes

¹ No doubt these radical-sounding thoughts are sufficiently titillating, and have sufficient resonance with many frustrating practical dilemmas and intractable moral disagreements that trouble us, that some people find in them enough appeal to be worth a defense. To most sober-minded people, however, the appeal of such thoughts does not last long, because the apparently exciting new vistas they appear to offer moral thinking turn out to be far less liberating than they at first seemed to be, once they have surveyed with even minimal care and seriousness. In any case, our inquiry here must begin with the recognition that Kant's reflection on morality begins with their resolute rejection.

² There are false negatives whenever we are dealing with a maxim (such as: "Give more to charity than the average person does") that does not violate universal moral laws, but could not itself be made into a law without contradiction. We can generate false positives by framing maxims that include enough specific information that the maxim would no longer have unwillable consequences if made into a universal law, but remains a morally objectionable maxim nonetheless – for instance, "If you are in need of money, borrow it from someone named Hilly Flitcraft on a Tuesday in August by promising to repay him, even though you have no intention of doing so." For a further discussion of these issues, see my book *Kant's Ethical Thought* (New York: Cambridge University Press, 1999), pp. 97-107. Self-appointed defenders of Kant ('self-appointed' because Kant never tries to use the universalizability test as a general moral criterion in the way they are trying to defend) will probably never abandon the noble, Grail-like quest for an interpretation of the universalizability test that enables it to serve this purpose, despite the history of miserable failure that has always attended this quest. I regard their attempts as worse than a waste of time, since they encourage critics of Kant's ethics to continue thinking, falsely, that something of importance for Kantian ethics turns on whether there is a universalizability test for maxims that could serve as such a general moral criterion.

³ John Stuart Mill, *Utilitarianism*, ed. George Sher (Indianapolis: Hackett, 1979), p. 24.

⁴ For documentation of this claim, see *Kant's Ethical Thought*, pp. 139-141.

⁵ It is the prior setting of an end as the condition of the imperative's validity for me that makes the imperative hypothetical in Kant's sense. "If you make a promise, keep it." is not a hypothetical imperative because the if-clause does not refer to an end that conditions the validity of the imperative. Likewise, categorical imperatives are categorical because their validity is not conditional on some end. A moral imperative may be conditional in other ways – for instance, there may be implied conditions that release us from a promise, in which case there is no categorical imperative at all to keep it under those conditions – but a valid moral imperative is always categorical in the sense that its rational validity does not depend on some prior setting of an end. The word 'prior' is crucial here, since categorical imperatives, in commanding us to act, also thereby always command us to set ends (according to Kant's theory, our own perfection and the happiness of others are the kinds of ends that are also duties). The thought that categorical imperatives command us to act without having any end at all is a nonsensical thought.

⁶ Since for Kant every action has an end to be produced, following a moral principle will always involve setting and achieving some end – for instance, fulfilling a promise will involve accomplishing the thing you promised to do. So it is just plain silly to represent Kantian ethics as caring nothing about the consequences of our actions – as is commonly done by those who don't understand the first thing about Kant's ethics, such as John Dewey (*Human Nature and Conduct* [1922] [New York: Random House, 1957], pp. 245-247). Nor do Kantian principles preclude using hypothetical reasoning about consequences – for instance, what would happen if anyone were permitted to make a promise without intending to keep it – from figuring in the reasoning that justifies the moral principle (ignoring this point frequently leads to a charge of inconsistency against Kant, as in Dewey, *op. cit.*, p. 246.) The point is rather only that the validity of a moral principle, such as 'Keep your promises' is not dependent on the actual achievement of any particular end to be produced by following the principle. Such criticisms of Kant are more often symptoms of an inconsistent procedure on the part of the critics. Starting from the mistaken idea that all practical reasoning

is instrumental in nature, and inferring from this that all moral reasoning must be justified by the particular consequences of the action, philosophers then see that there are clear counterexamples to this consequence. So they try to save their original dogma by appealing not to actual consequences but to the expected consequences of a principle's being generally followed or to the imagined consequences of its being followed or not followed under certain ideal counterfactual circumstances (in other words, using the same kind of reasoning that Kant uses). They then conclude that maybe there is something right in Kant's theory after all (see Dewey, *op. cit.*, pp. 246-247); but they erroneously regard it as consistent with their own dogma that all practical reasoning is instrumental and oriented toward actual consequences, and inconsistent with Kantian principles. But the incoherence is in their views, not in Kant's.

⁷ This triad represents the three conditions Kant places on concept formation, with the third condition applying only to the concepts of individuals. "Form" refers to the kind of generality involved in a concept, created by the understanding according to the judgment-forms and categories; "matter," to the intuitive content or possibility of providing a sensible object for a concept. "Complete determination" means that for every pair of contradictory predicates, one and only one of them belongs to the concept. When a concept is completely determined, it is (according to Leibnizian doctrine) the concept of an individual rather than a universal concept. Why does Kant choose this triad, drawn from his logic of concepts, to systematize the formulas for the supreme principle of morality? I have attempted to answer this question in "The Moral Law as a System of Formulas", in H. Stolzenberg and H. F. Fulda (eds.) *Architektur und System in der Philosophie Kants*. Hamburg: Felix Meiner Verlag, 2001.

⁸ Schopenhauer, *The World as Will and Representation*, tr. E.J.F. Payne (New York: Dover, 1958) 1:514-526, *On the Basis of Morality*, tr. E.J.F. Payne (Indianapolis: Bobbs-Merrill, 1965), pp. 53-55.

⁹ Kant himself may have made it difficult for readers to see this point, by insisting in the Critique of Practical Reason that morality must countenance only 'formal principles' and eschew 'material principles' that presuppose an end (*Practical Reason* 5:27). But here he is using 'end' and 'material principles' in the traditional sense only. It should be appreciated that already in the *Groundwork*, he distinguished 'formal principles' from 'material principles' precisely in terms of the kind of motive (or end) on which they were grounded (*Groundwork* 4:427-428). A formal principle is *never* for Kant a principle that is not grounded on any end as its motive. Here the overemphasis on FUL and FLN again does mischief, by persuading people that we may identify a 'formal' principle simply using the universalizability tests, without recourse to any substantive value or end to serve as the motive for following such a principle. A careful reading of *Groundwork* 4:427-429 reveals this reading of Kant to be quite mistaken.

¹⁰ For a fuller account of my interpretation of this argument, see *Kant's Ethical Thought*, pp. 124-132. A similar interpretation has been defended by Christine Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), pp. 106-133.

¹¹ Why not also our own happiness and the perfection of others? We honor rational nature as an end in itself by making it our end to increase its capacity rationally to set and pursue ends, and the general name for this capacity is 'perfection'. We honor it also by making our end the ends set through its exercise, and the name for the totality of ends a rational being proposes for itself is its 'happiness'. We honor rational nature in others only to the extent that we further the perfections in themselves that they also set as ends, so morality bids us to pursue their perfection only insofar as it falls under the heading of their happiness. We need no moral constraint to pursue our own happiness except insofar as we are tempted to make ourselves less perfect by neglecting it, so morality bids us pursue our own happiness only insofar as it falls under the heading of our perfection.

¹² The first of these translations is Lewis White Beck's (*Foundations of the Metaphysics of Morals* [Indianapolis: Bobbs-Merrill, 1959], p. 54), the second is Mary Gregor's (Kant, *Practical Philosophy* [New York: Cambridge University Press, 1996], p. 85). My own translation (*Groundwork of the Metaphysics of Morals* [New Haven: Yale University Press, 2002], p. 54) is also a bit less literal than the one I have just presented: "one of which unites the other two in itself". But it does capture the feature of the passage to

which I am calling attention. It is true that the original text can be read so that it does not positively exclude the possibility that each of the three formulas unites the other two. If I say of the three musketeers: “One of them would give his life for the other two,” it is natural to understand me to be saying not only of one specific musketeer but of *each* of the three, Athos, of Porthos and Aramis, that he would give his life for his two comrades. But this is only because we have no reason to single out any one of them as more self-sacrificing than the other two, and so we have reason to treat “one” in this context as meaning “each one”. However, in the *Groundwork* Kant has already singled out FA by saying that it follows from FUL and FH, but he has not made any comparable claim about either FUL or FH. So it is not natural to read “one” (die eine) here as if it were equivalent to “each” (*jede* or *irgend eine*). If Kant had meant *jede* or *irgend eine*, he could have said so. The (mis)reading of the passage is normally used to suggest, at the outset, a kind of equality of status between the three formulas, but usually this is nothing but a front for the common reading which privileges FUL over the other two formulas. That in turn usually goes along with treating FUL as a universal moral criterion, or interpreting it as a procedure for “constructing” the entire content of ethics, or a lot of other false and philosophically indefensible notions that fundamentally misunderstand Kant’s moral philosophy.

¹³ For more on Kant’s views about friendship and religion, see *Kant’s Ethical Thought*, pp. 274-282, 309-320. The spirit of Kantian ethics has often been characterized as ‘individualistic’, on account of the priority Kant gives to the value of autonomy or self-legislation, to individual rights and freedom, to thinking for oneself, and because Kant regards only individuals, never groups of people, as bearing moral responsibility (though – what is seldom appreciated -- he does regard both the cause and the cure for moral evil as social). Yet once we see that the fundamental principle of morality, formulated as FRE, gives absolute priority to achieving a community, that is, a convergence or consilience, among all the ends of all rational beings, we should also recognize that the spirit of Kantian ethics is, at a very fundamental level, exactly the reverse of individualistic.

¹⁴ Examples are H. J. Paton, *The Categorical Imperative* (London: Hutcheson, 1947), p. 130; Onora O’Neill, *Constructions of Reason* (Cambridge: Cambridge University Press, 1989), p. 127; Paul Guyer, “The Possibility of a Categorical Imperative,” in Guyer (ed.) *Groundwork of the Metaphysics of Morals: Critical Essays* (Lanham, MD: Rowman and Littlefield, 1998), p. 216.

¹⁵ Klaus Reich, “Kant and Greek Ethics II,” *Mind* 48 (1939), pp. 452-453. Cited below as “Reich” by page number.

¹⁶ See previous note, also A. R. C. Duncan, *Practical Reason and Morality* (London: Nelson, 1957), pp. 175-182, cited below as “Duncan”. See also H. J. Paton, “The Aim and Structure of Kant’s *Grundlegung*,” *Philosophical Quarterly* 8 (1958), pp. 112-130 and Brendan E. A. Liddell, *Kant on the Foundation of Morality* (Bloomington: Indiana University Press, 1970), pp. 138-140.

¹⁷ Cicero, *On Duties*, cited by book and section. Cicero, *De officiis*, Scriptorum, classicorum bibliotheca Oxoniensis (Oxford: Clarendon Press, 1994). For a recent English translation, see Cicero, *On Obligations*, translated by P. G. Walsh (Oxford: Oxford University Press, 2000).

¹⁸ Reich relies on Garve’s account of Cicero’s intention, which is that Cicero was trying to “resolve the apparent conflict between duty and interest” (Reich, p. 455). But this is surely misleading, for it suggests that it was Cicero’s aim to devise strategies for making the (apparent) conflict disappear, whereas the plain import of *On Duties* is that there is no such conflict, and it is only human error or vice that leads people to think there is. Of course, Kant is even less interested than Cicero in the project Garve describes. For it is his view that when people take steps to reduce the “apparent” conflict of duty and interest (which in Kant’s view is sometimes not apparent but quite real), they do so mainly by deceiving themselves about what duty demands, and softening these demands so that they do not infringe on self-interest. Kant is as far as it is possible for anyone to be from wanting to help them to do this.

¹⁹ G. E. M. Anscombe, "Modern Moral Philosophy," *Philosophy* 33 (1958), reprinted in Roger Crisp and Michael Slote (eds.), *Virtue Ethics* (Oxford: Oxford University Press, 1997), cited in the latter source, by page number, as "Anscombe".

²⁰ Of course he does not accept – and has no reason to accept – Anscombe's contention that a "law conception of ethics" makes sense only in a framework of thought where a divine legislator is thought of as the source of moral obligation; still less does he accept Anscombe's contention that "the idea of 'legislating to oneself' is absurd," because "the concept of legislation requires a superior power in the legislator" (Anscombe, p. 27). We find as far back as Socrates the idea that the most important ruler-ruled relationship is the relationship to oneself (Plato, *Gorgias* 491d5-10); the idea of constraint through laws and principles makes just as much sense when the constraint is conceived as self-constraint through reasons as when it is conceived as coercive constraint through external force. Simply to assume the opposite without argument is simply to dismiss out of hand, and for no good reason, the most basic idea of Kantian ethics.

²¹ For example, Paul Guyer, "Kant's Deduction of Principles of Right," Mark Timmons (ed.), *Kant's Metaphysics of Morals: Interpretive Essays* (Oxford: Oxford University Press, 2002), pp. 23-64; Bernd Ludwig, "Whence Public Right? The role of Theoretical and Practical Reasoning in Kant's Doctrine of Right," Timmons (ed.), *Kant's Metaphysics of Morals*, op. cit., pp. 159-184. The other side of this interpretive dispute, the one apparently supported by Kant's criticism of Cicero, is defended in this same volume by me, "The Final Form of Kant's Practical Philosophy," pp. 5-10, by Markus Willascheck, "Which Imperatives for Right? On the Non-Prescriptive Character of Juridical Laws in Kant's Metaphysics of Morals," pp. 65-88 and by Thomas Pogge, "Is Kant's *Rechtslehre* a 'Comprehensive Liberalism'?", pp. 133-158. There is some connection, but by no means an identity, between the issue being debated here and the older controversy in the German literature between defenders of what has been called the 'independence thesis' (such as Julius Ebbinghaus, Klaus Reich and Georg Geismann) and critics of it (such as Wolfgang Kersting and Bernd Ludwig). A good discussion of this issue is found in Pogge, loc. cit., pp. 150-151.

²² The emphasis on truth is especially identified with Wollaston, whose views were prominently criticized (though not using his name) in Hume's *Treatise on Human Nature*, edited by L. A. Selby-Bigge (Oxford: Clarendon Press, 1967), p. 461. It is even a bit odd that Kant should identify this view with Cumberland (even though it is stated prominently in the opening chapter of *De Legibus Naturae* [1672]), since Cumberland is more often thought of as an ethical eudaimonist. The probable explanation is that Cumberland is the only one of these authors who wrote in Latin, and Kant did not read English.

²³ In the *Metaphysics of Morals*, Kant criticizes Aristotle's principle not on the ground that it is analytic, but on the ground that it is *false*, since (he argues) it gives a false account of what virtue is (*Morals* 6:404).

²⁴ The fact that some names appear more than once (when the *Lecture* versions of Kant on the history of ethics are taken into account) should not disturb us. It is quite possible to interpret Epicurus, for instance, as a representative of the ancient "ideal" conception of ethics and also to realize that his views can be (and are) appropriated by moderns to support a principle of hedonistic eudemonism. In some philosophers, such as Mandeville and Baumgarten, it is easy enough to find endorsements of both of the views with which Kant associates their respective names. (If there is an inconsistency here, the blame lies at their door rather than at Kant's.) And it is quite possible to interpret the British rationalist principle of truth either as an analytic claim or (more sympathetically) as a version of rational perfectionism.

²⁵ The problem is that there are just too many ways in which we could come to desire things we have no reason to desire, be disposed to approve of things we have no reason to approve, and take as reasons things that aren't genuine reasons at all. I will admit, or rather earnestly maintain, that our having a natural desire for something is good evidence that we have a reason to desire it, and even that it is good. (Thus J. S. Mill's much maligned argument to that effect in Chapter 4 of *Utilitarianism* seems to me a sound argument.) But then we have to suppose that there are good things, and genuine reasons for desiring or approving of good things, and that our natural dispositions somehow involve our being in contact with those reasons. The

problem with empiricist theories is that when presented with this obvious point, the empiricists won't admit it. When you ask them to tell you what reasons we have for doing anything, and insist that they give you some answer that at least has the general form of being a genuine reason, they get all nervous and huffy and accuse you of metaphysics, obscurantism, and god knows what other misdemeanors. Accepting their theories apparently requires us either to stop asking for reasons at all or else to allow to count as reasons things that are transparently not genuine reasons at all. Such defensiveness is a sure sign that there is something deeply wrong with their position, and that at some level, they know it.

²⁶ Of course it is reasonable to ask the rationalist what these properties are, whether they are natural or non-natural properties, how we can know about them, and so forth. And it may be that their theories about these matters fall far short of being satisfactory. But at least their account of what reasons are has the virtue that what it says are reasons might actually be reasons. (That an action is right or good or ought to be done is a reason for doing it, whereas the fact that we are disposed to desire something is transparently not a reason for desiring it and that we are disposed to approve of something is transparently not a reason for approving it.) Some Kantians who like to call themselves 'constructivists' think either that Kant has a superior theory to the rationalists on these points, or else at any rate that inspired by Kant, they have devised a superior theory. I think they are fooling themselves about this. Kant was perhaps a constructivist about mathematics in some intelligible sense, but no intelligible sense has yet been given to the term "Kantian constructivism" in ethics. As far as I can see, Kant has no better metaethical theory than the rationalists do, perhaps because he was interested only marginally, if at all, in the questions such theories are designed to answer.

²⁷ Many of the ways that British moralists in the rationalist tradition anticipated Kant are explored in Stephen Darwall's excellent book, *The British Moralists and the Internal 'Ought' 1640-1740* (New York: Cambridge University Press, 1995).