Deriving the Supreme Principle of Morality from Common Moral Ideas

In the Preface to the *Groundwork of the Metaphysics of Morals*, Kant sets out his goals: to locate and to establish the supreme principle of morality (GMS 392). He devotes Section I to the first goal. Working under the assumption that there is a supreme principle of morality, he tries to locate it in the sense of specifying its content.¹ Kant strives to find the supreme principle that, on reflection, we hold to be at work in our moral practice. His attempt rests on appeals to (what he takes to be) ordinary moral reasoning. Near the end of *Groundwork* I, he proclaims success. “Considering the moral knowledge of common human reason we have thus arrived at its principle, a principle it admittedly does not think about abstractly in such a universal formulation; but which it really does always have in view and employs as the standard in its judging” (GMS 403). This principle is the categorical imperative. In *Groundwork* I, Kant’s main concern is to show that if there is a supreme principle of morality, then it is this imperative. It is not until *Groundwork* III that Kant tries to establish the categorical imperative, that is, to prove that we are all rationally compelled to conform to it.

This paper focuses on Kant’s attempt to locate the categorical imperative—an attempt which, in the idiom of contemporary Kant scholarship, is called his “derivation”

¹ Kant’s main task in *Groundwork* II also seems to be to derive the supreme principle of morality—in all the complexity of its various formulas.
of this principle. In *Groundwork* I Kant derives one particular formulation of the
categorical imperative, namely a version of his famous Formula of Universal Law: *act
only on that maxim through which you can at the same time will that it become a
universal law.* (GMS 420-421).\(^2\) The title of *Groundwork* I is “Passage from the
Common Rational Knowledge of Morality to the Philosophical.” Kant’s starting point is
“common rational knowledge of morality,” which is a fancy term for common moral
ideas. How, from this starting point, does Kant derive the Formula of Universal Law?
This paper sketches an answer to this question.

In my view, Kant’s *Groundwork* I derivation of the Formula of Universal Law
takes place in three main steps. First, Kant tries to pinpoint criteria that we, on reflection,
believe that the supreme principle of morality must fulfill. Second, Kant attempts to
establish that no possible rival to the Formula of Universal Law fulfills all of these
criteria. Third, at least implicitly Kant argues that the Formula of Universal Law remains
as a viable candidate for a principle that fulfills all of them. With these three steps, Kant
strives to prove that if there is a supreme principle of morality, then it is this formula. In
short, Kant argues by elimination. When we have before us a clear notion of the
characteristics the supreme principle of morality must possess, Kant suggests, we are able
to eliminate every candidate for this principle except the Formula of Universal Law (or
equivalent principles). I call this interpretation of Kant’s derivation the “criterial

---

\(^2\) Strictly speaking, the principle is a preliminary version of the Formula of Universal
Law, namely: “I ought never to act in such a way that I could not also will that my maxim
should become a universal law” (GMS 402).
reading,” since it emphasizes that Kant develops criteria that any viable candidate for the supreme principle of morality must fulfill.

The criterial reading of Kant’s derivation is not the only one philosophers have proposed. (For alternatives see, for example, Aune, 1979 and Korsgaard, 1996). Elsewhere, I have tried to demonstrate that the criterial reading compares well with other interpretations in terms of its textual plausibility and philosophical fruitfulness (Kerstein, 2002). Here I limit myself to illustrating how the criterial reading helps us make sense of *Groundwork* I. In particular, I focus on how Kant appeals to “common rational knowledge of morality” in order to develop criteria for the supreme principle of morality.

The *Groundwork* I derivation culminates in the introduction of the Formula of Universal Law. In the sentence that immediately precedes its introduction, Kant poses a question: “But what kind of law can that be, the idea of which must determine the will, even without considering the expected result, if that will is to be called good absolutely and without qualification?” (GMS 402). Kant is in effect asking which principle (law) can possess certain characteristics, namely ones that, according to ordinary moral thinking, the supreme principle of morality must possess. These characteristics are crystallized in the famous three “propositions” that Kant tries to establish in *Groundwork* I. The propositions encapsulate criteria for the supreme principle of morality. So by tracing how Kant arrives at his propositions, we will gain an understanding of how he develops these criteria.

The paper unfolds as follows. Parts II-III concern Kant’s three propositions and the criteria for the supreme principle of morality implicit in them. The bulk of the discussion focuses on Kant’s first proposition, namely that an action has moral worth just
in case it is done from duty. Kant relies far more directly on ordinary moral thinking in
defending this proposition than he does in defending the other two. Part IV illustrates
how Kant might use criteria he develops in order to show that rivals to the Formula of
Universal Law cannot be the supreme principle of morality. Unfortunately, these criteria
do not enable him even to come close to eliminating all rivals. But Kant’s argument gains
in strength when we recognize that he suggests an additional criterion for the supreme
principle of morality: it must be such that a plausible set of moral prescriptions (i.e.,
plausible relative to common rational moral cognition) would stem from the principle.
Part V focuses on this additional criterion. According to the criterial reading, the final
step of the derivation is to show that the Formula of Universal Law remains as a viable
candidate for a principle that fulfills all of the criteria, including that it generate a
plausible set of moral prescriptions. Determining whether the Formula of Universal Law
fulfills this criterion would obviously require delving into the details of how to interpret
this principle. That is a project for another occasion. The final section (VI) of this paper is
devoted to a concern one might have regarding the interpretation of the derivation offered
here. The interpretation emphasizes the extent to which it depends on appeals to ordinary
moral thinking. But how are such appeals, which seem to be appeals to experience, to be
reconciled with Kant’s view that the supreme principle of morality is an a priori
principle?

Before focusing on *Groundwork* I and the criteria for the supreme principle of
morality that Kant there suggests, it will be helpful to consider what he means by
“supreme principle of morality” in the first place.

I.
According to (what I call) Kant’s basic concept, the supreme principle of morality would possess four characteristics. It would be practical, absolutely necessary, binding on all rational agents, and would serve as the supreme norm for the moral evaluation of action. I call this concept of the supreme principle of morality basic because Kant suggests it in the *Groundwork’s* Preface.

To say that a principle must be the supreme norm for the moral assessment of action suggests several things. The principle would obviously distinguish between morally permissible and morally impermissible actions as well as specify which actions are morally required. In addition, whether an action was morally good would depend on how it related to this principle. Kant implies, for example, that no action that violated the principle would count as morally valuable (GMS 390). Finally, as the supreme norm for the moral assessment of action, the supreme principle of morality would be such that all genuine duties would ultimately be derived from it (see GMS 421; see also GMS 424, 425). The supreme principle would justify these duties’ status as such.

Kant says that the supreme principle of morality “must be valid not merely for human beings, but for all *rational beings as such*” (GMS 408; see also GMS 389, KpV 32). The supreme principle of morality would have an extremely wide scope: one that extended not only to all rational human beings, but to all other rational beings, if any others exist, for example, to God, angels, and intelligent extraterrestrials.

A third feature the supreme principle of morality would have to possess is “absolute necessity” (GMS 389). On every agent within its scope, for Kant every rational agent, the principle would hold without exception (GMS 408). For us, human agents, the supreme principle of morality would be an unconditional command (i.e., a categorical
imperative in one sense of the term). That we were obligated to perform the action it specified would not be conditional on our having any particular set of desires.

Finally, for Kant the supreme principle of morality must be practical, that is, a rule on account of which agents can act. Kant implies this in the *Groundwork* Preface by specifying that morally good actions involve an agent’s acting for the sake of the moral law, that is, the supreme principle of morality (GMS 390). For Kant the supreme principle must be able to figure directly in an agent’s practical deliberations.

Kant claims to find in ordinary moral thinking agreement that the supreme principle of morality would have to have at least some of these features. He suggests, for example, that, according to the “common idea of duty,” a moral law “must carry with it absolute necessity” (GMS 389).

II.

Now that we have an idea of what, according to Kant, the supreme principle of morality would have to be like, let us focus on the details of his *Groundwork* I attempt to show that if there is such a principle, then it is the Formula of Universal Law. As I indicated, Kant’s attempt involves his developing three propositions, each of which implies a criterion for the supreme principle of morality. It makes sense to begin with Kant’s first proposition. Although Kant does not explicitly state it, it is widely, and, I believe, correctly taken to be the following: An action has moral worth if and only if it is done from duty.

As a first step towards understanding this proposition we need to delve briefly into Kant’s famous discussion of a good will. For Kant tells us that the concept of duty “includes that of a good will, the latter however being here under certain subjective
limitations and obstacles” (GMS 397). Let us focus on Kant’s discussion of the good will as it relates to us, agents who can indulge their inclinations and thereby act contrary to what morality requires. In this context, Kant seems to use the notion of a good will in two ways. According to the first usage, a good will is a particular sort of willing or, what for him amounts to the same thing, of acting. Kant writes of “the unqualified [uneingeschränkten] worth” of actions (GMS 411). These are presumably actions done from duty; for he has earlier stated that actions from duty have “unconditional” and “moral” worth (GMS 400). Since, according to Kant, the good will is good without qualification [ohne Einschränkung], it appears that sometimes “good will” refers to a certain kind of action, that is, action done from duty.

According to a second usage of “good will,” it refers not to a particular kind of action an agent might perform but rather to a kind of character she might have. An agent has a good will on this usage, I believe, just in case she is committed to doing what duty requires, not just in this or that particular action, but overall. If an agent has this commitment, then she will presumably sometimes act from duty. (For example, she will invoke duty as her incentive to do what is morally required in cases in which she is tempted by her inclinations to act contrary to what morality demands.) Kant intimates that having a good will amounts to having a certain kind of character in the first paragraph of *Groundwork* I (GMS 393; see also GMS 398-399).

Kant’s “particular action” conception of a good will is more important for our purposes. Kant suggests that a good will in the sense of good willing is equivalent to acting from duty. Moreover, according to “common reason” (GMS 394) this willing (or, equivalently, acting) has a special, moral worth. First, it is unconditionally good. In all
possible circumstances in which it appears, a good will is not only good, but also has the same level of goodness. Even if a good will “were completely powerless to carry out its aims; if with even its utmost effort it still accomplished nothing, so that only good will itself remained . . . even then it would still, like a jewel, glisten in its own right, as something that had its full worth in itself” (GMS 394). Second, according to ordinary moral thinking, the worth of a good will is especially high, Kant claims. We take a good will (good willing) to be preeminently valuable (see GMS 394 and 401). That presumably implies that, in our view, no particular action that is not done from duty is as valuable as any action that is done from duty.

Let us now return to Kant’s first proposition. It states that an action has moral worth if and only if it is done from duty or, equivalently, that all and only actions done from duty have moral worth. The two key concepts in this proposition are obviously those of moral worth and of acting from duty. Moral worth, as we just noted, is unconditional and preeminent worth. At this stage in his argument, Kant does not explain precisely what acting from duty amounts to. But from the *Groundwork*’s Preface, it’s easy to discern the basic idea he has in mind and takes his reader to have in mind as well. Acting from duty is doing something “for the sake of” the moral law (GMS 390). In other words, to act from duty is to do something because a valid moral principle (or at least a principle one takes to be valid) prescribes that one do it. A more rigorous account of acting from duty emerges from Kant’s discussion of his third proposition.

In the *Groundwork*, Kant apparently finds it unnecessary to argue that all actions done from duty possess moral worth. Consider, for example, his discussion of self-preservation. Kant suggests that we have a duty to preserve our life and that, the vast
majority of the time, when we take steps to preserve it we are acting from an immediate inclination to stay alive. “But for that reason,” Kant says, “the often-fearful care that most people take for their lives has no intrinsic worth, and the maxim of their action has no moral merit. They do protect their lives in conformity with duty, but not from duty” (GMS 398). Kant takes it to be obvious that if a person preserves his life not from inclination but from duty, his maxim has moral merit, and thus acting on it has moral worth. He assumes that ordinary moral thinking needs no coaxing in order to see that actions done from duty possess moral worth.

In contrast, Kant does think we need a bit of help in order to discern that only actions from duty have moral worth. He highlights two conditions on actions with such worth, both of which he takes to be accepted by common rational moral cognition. He then intimates that no action from inclination could meet these conditions.

Kant introduces the first condition in the *Groundwork* Preface:

in the case of what is to be morally good, it is not enough that it conform with the moral law; but it must also be done *for the sake of the law*; without this, that conformity is only very contingent and precarious, since a ground that is not moral will indeed now and then produce actions in conformity with the law, but it will also often produce actions contrary to the law. (GMS 390)

Morally valuable action, Kant here suggests, is that action done from a motive that will not produce actions contrary to duty. Kant maintains that acting “for the sake of the law,” that is, doing something because you take it to be required by moral principle, meets this condition, while acting from inclination does not.
Kant invokes this condition in his famous discussion of the “philanthropist” (or “humanitarian”) (GMS 398). The philanthropist has an inclination to promote the well-being of others. But, according to Kant, acting from this inclination is like acting from the inclination to pursue honor, “which, if fortunate enough to aim at something generally useful and consistent with duty, something consequently honorable, deserves praise and encouragement but not esteem” (GMS 398, see also Rel 30-31). Here Kant underscores the possibility that in acting from an inclination to help others, that is, from sympathy, an agent might do something that conflicts with duty. To echo a well-known example (Herman 1993: 4-5), someone might, because of his sympathetic temperament, act on an inclination to help someone he sees late one night hurriedly struggling to move a sculpture out the back door of an art museum and into his waiting car. Since the philanthropist is acting from an inclination and thereby doing something that might fail to accord with duty, his action, Kant suggests, does not have moral worth.

In his discussion of the philanthropist Kant points to a further condition he places on an action’s having moral worth (Herman 1993: 5-6). Kant says that the maxim on which the philanthropist acts “lacks the moral merit” of the sort of maxim an agent acts on when acting from duty (GMS 398). Kant does not tell us explicitly what the philanthropist’s maxim is. From the description Kant provides, however, we can assume that it is something like the following: “Because I want to help others, I will promote their happiness.” This maxim, says Kant, lacks moral merit, and it is not hard to pinpoint a reason why. The maxim reflects no commitment to the action’s being morally permissible, that is, in accordance with what moral principle requires. According to ordinary moral understanding, the grounds of a morally valuable action, that is, its
motive, must express an interest in the action’s moral rightness, intimates Kant. This is the second condition that any action having moral worth must fulfill. Actions done from duty fulfill both this condition and the previous one, while actions form inclination do not, holds Kant.

Kant’s first proposition and his defense of it have attracted ample critical attention. Kant is perhaps too quick to conclude that, according to ordinary moral thinking, an action has moral worth only if it fulfills his two conditions (and thus only if it is done from duty). He might also be precipitous in assuming widespread endorsement of the notion that all actions from duty have moral worth. My own view (Kerstein 2002: 114-138) is that Kant is on much stronger ground in claiming that, according to ordinary moral thinking, all actions from duty have moral worth than he is in claiming that only actions from duty have moral worth. In any case, Kant’s main appeals to ordinary moral thinking occur in his discussion of the special value possessed by a good will, as well as in the closely related discussion leading up to his first proposition.

III

The arguments Kant suggests for the second and third propositions are far less directly tied to intuitive moral judgments than his arguments for the first. In his “second proposition,” Kant says that “the moral worth of an action done from duty has its moral worth, not in the objective to be reached by that action, but in the maxim in accordance with which the action is decided upon; it depends, therefore, not on actualizing the object of the action, but solely on the principle of volition in accordance with which the action was done . . .” (GMS 399-400). Later Kant says that the moral worth of an action does not depend “on the result expected from that action” (GMS 402; see also GMS 435).
Kant here invokes the notion of a principle of volition or maxim. We’ve already made use of this notion, but it makes sense to pause here to get a more precise idea of what a maxim is. The brief account of maxims that follows is certainly not the only plausible one, but it will serve to fix ideas. A maxim is a “subjective principle of acting” (GMS 421, note; see also GMS 400, note). It is a subjective principle in that it is held by some agent, it can be freely adopted or discarded by her, and it applies only to her own actions. An agent’s maxims are principles of acting in that they play a role in the generation of her actions. An agent acts on maxims. When fully specified, a maxim includes a description of a kind of action to be performed in a kind of situation, as well as a specification of the agent’s end and his incentive in performing it. An example of a fully-specified maxim is the following: “From self-love, during my free time I exercise in order to stay in shape.” (Self-love is the agent’s incentive; staying in shape is her end.) According to Kant, whenever an agent acts, she does so on some maxim, even though she might not have it explicitly in view.

Kant’s second proposition says essentially that an action done from duty derives its moral worth from its maxim rather than from its effects. The proposition relies on a distinction between an action, which is always done on some maxim, and its effects. For Kant, to act is to exercise one’s will (Kerstein 2002: 20-21). It is to try, based on some principle of volition, to realize a state of affairs (an object or end). This state of affairs, or whatever state of affairs actually results from the action, is an effect of the willing. Acting consists in the willing itself, not in its effects (see GMS 400). According to the second proposition, it is the maxim behind an action done from duty that gives it moral value, rather than the action’s results.
Implicit in *Groundwork I* is a straightforward argument for the second proposition. Suppose that, contrary to it, the moral worth of an action from duty *did* stem from its effects. There would, then, be possible circumstances in which an action from duty did not have moral worth, namely ones in which the action failed to produce certain effects. For Kant, however, if an action is done from duty, then it has moral worth, no matter what the circumstances may be. His first criterion incorporates this view. Moral worth is “unconditional,” Kant suggests (GMS 400). Therefore, as the second proposition indicates, the moral worth of an action from duty does *not* stem from its effects.

According to Kant’s third proposition, “duty is the necessity of an action from respect for law” (GMS 400, emphasis omitted). This proposition fills in some details regarding what it means for an action to be done from duty. According to the proposition, if an action is done from duty, then what determines it is “objectively the law and subjectively pure respect for this practical law” (GMS 400-401). By “law”, Kant means a universally binding and absolutely necessary practical principle. When an agent acts from duty, Kant here implies, his action stems from the notion, which is incorporated into his maxim, that a practical law requires it. Kant even says that an action from duty “is to exclude totally the influence of inclination” (GMS 400). So, in his view, an agent who needs to rely on an inclination in order to get something done fails to act from duty. If an agent acts from duty, the notion that a law requires her action itself generates enough motivation for her to do it. It generates this motivation, Kant suggests, at least in part by producing in her a feeling of respect for the law. Kant develops his concept of respect in detail in the *Critique of Practical Reason* (see KpV 71-89). It is very complex, and we have no need to explore it here. But we do need to hold in view that, according to Kant’s
third proposition, when an agent acts from duty, her notion that her action is required by a practical law provides her with sufficient motive for doing it. In other words, this notion gives her a ground sufficient to determine her will.

But how does Kant defend this proposition? He suggests, but does not explicitly make, the following argument (see GMS 401). Suppose that in an action done from duty the notion that the action was required by a practical law did not give an agent sufficient motive to perform it. In that case, Kant suggests, the additional motive necessary for the agent to perform the action would have to be the agent’s expectation that her action would bring about certain effects (GMS 401). But now further suppose that the action did not produce the expected effects. In that case, the agent would be rationally compelled to agree that the action had less value than it would have had if the expected effects had come to fruition. After all, if, in the agent’s view, the action’s value was not at all contingent on the effects being produced, then why would she need to acquire part of her motivation for doing it from the prospect that the effects would be produced? But if an action done from duty has less value than it otherwise would have as a result of its not producing certain effects, then its value is not unconditional. And this result conflicts with Kant’s first proposition, according to which all actions have moral worth, that is, the same (high) degree of value in every possible situation in which they occur. The result also conflicts with his second proposition, since according to it the moral worth of an action does not depend (at all) on the action’s effects. So it makes sense for Kant to suggest, as he does (GMS 400), that his third proposition follows from the previous two.

Kant’s main aim in articulating his three propositions is to derive the supreme principle of morality, that is, to show that if there is such a principle, then it is the
Formula of Universal Law. Each one of the propositions implies a corresponding
criterion that the supreme principle of morality must fulfill. According to the criterion
implicit in the first one, the supreme principle of morality must be such that all and only
actions conforming to it because the principle requires it, that is, all and only actions done
from duty, have moral worth. The second proposition implies that whatever the supreme
principle of morality is, the moral worth of conforming from duty to it must stem from
the maxim of the action, not from its effects. According to the criterion implicit in the
third proposition, the supreme principle of morality must be such that an agent’s notion
that it is a practical law and that it requires her to do something gives her sufficient
motive to do it.

Let us again note that in the sentence preceding his initial presentation of the
categorical imperative, Kant asks: “But what kind of law can it be, the idea of which must
determine the will, even without considering the expected result, if that will is to be
called good absolutely and without qualification?” (GMS 402) He is, in effect, asking
what law (principle) can fulfill each of these three criteria for the supreme principle of
morality: the third criterion, which invokes an agent’s representation of a law as a
sufficient motive for her action; then the second criterion, which incorporates the notion
that the moral worth of an action does not stem from its effects; and finally the first
criterion, which specifies when an action, that is, an instance of willing, has moral and
thus unconditional worth. If we can show that a particular principle is unable to fulfill any
one of these criteria, then we can, Kant suggests, eliminate it as a viable candidate for the
supreme principle of morality. If Kant's derivation of the categorical imperative is
successful, then we should be able to see that the only principle that remains as a viable
candidate for satisfying all three of these criteria (plus those implicit in Kant's basic concept of the supreme principle of morality) is the Formula of Universal Law (or an equivalent principle).

IV

Unfortunately, it would be unduly optimistic to say that we are in position to see this. Kant moves extremely quickly from the criteria he develops to the conclusion that the only viable candidate for fulfilling all of them is the Formula of Universal Law. He seems to leave it to us to fill in the details regarding precisely how rivals get eliminated. Even if, employing Kant's criteria, we eliminate all rivals that come to mind, it is not clear how we can be confident that we have not overlooked some other rival (see Kerstein 2002: 140-144). Nevertheless, we are well-situated to see how we might use Kant’s criteria to dismiss some well-known principles as viable candidates for the supreme principle of morality.

Kant does not explicitly argue against utilitarianism. But let us consider a utilitarian principle, U: ‘Always perform a right action: one that yields just as great a sum total of well-being as would any alternative action available to you.’ Let us suppose, as it seems reasonable to do, that the utilitarian embraces this principle largely on the grounds of her being convinced of the following. First, the amount of goodness in the world depends solely on the sum total of individual well-being in it--the higher the sum total, the more goodness. Second, the rightness of an action depends solely on the goodness of its consequences. More precisely, an action is right just in case that which results from it is at least as good as that which would have resulted from each of the alternative actions available to the agent.
Although U derives from these unKantian convictions, it would be precipitous to dismiss it as a candidate for the supreme principle of morality on the grounds of a manifest failure to conform to Kant’s basic concept of this principle. U could, it seems, be a practical, absolutely necessary, universally binding, fundamental norm for moral evaluation of action.

But U runs afoul of Kant's further criteria for the supreme principle of morality. The utilitarian might insist that an agent can, from duty, comply with U. After all, what would prevent her from performing a right action just because U commands her to do so? Yet she is committed to the following view: whether an agent’s conforming to U from duty has moral worth depends solely on the action’s effects, specifically its effects on well-being. For she holds that the amount of goodness in the world (including the “moral worth” of actions) depends *solely* on the amount of well-being in it. So the utilitarian cannot, rationally speaking, maintain that U fulfills Kant’s second criterion, namely that the supreme principle of morality be such that the moral worth of conforming to it from duty stems not at all from that action’s effects.

If we think of a consequentialist principle as one according to which the goodness of any action depends to some extent on the action’s effects (in addition to the “effect” that the action has taken place), then it is easy to show that no consequentialist principle fulfills Kant's second criterion. For even the staunchest proponent of such a principle would have to acknowledge that he is committed to the view that the value of acting from duty depends at least in part on what that action produces.

Based ultimately on an appeal to the notion that, according to ordinary moral thinking, actions from duty have a special worth, Kant develops three criteria for the
supreme principle of morality. Assuming these criteria are sound, Kant has solid grounds for dismissing some of the categorical imperative's rivals for status as viable candidates for the supreme principle of morality.

V

However, if, as a basis for dismissing rivals, Kant has only these three criteria, coupled with those implicit in his basic concept of the supreme principle of morality, he is vulnerable to a serious criticism. Using these criteria, he would be helpless to eliminate rivals that, one would think, would have almost no chance of being the supreme principle of morality. Consider the bizarre principle, BP: “Act only on that maxim such that you cannot, at the same time, will that it become a universal law.” Assuming that the categorical imperative could be a universally valid, absolutely necessary, supreme practical principle, why couldn’t BP be such a principle? What argument does Kant have at his disposal that would show it to be impossible for BP to have these characteristics? Moreover, it seems that a proponent of BP would be able consistently to maintain that an action has moral worth if and only if it is done because BP requires it, that such an action’s moral worth would not stem from its effects, and so forth. He would not be rationally compelled to acknowledge that BP runs afoul of the criteria implicit in Kant's three propositions.

Another, less provocative, example of a principle Kant would be unable to dismiss on the basis of his criteria is the following principle of weak universalization, WU: “Act only on that maxim which, when generalized, could be a universal law.” WU is not equivalent to the Formula of Universal Law. Kant himself suggests that a maxim of non-beneficence could, when generalized, constitute a universal law (GMS 423). Since a
world where no one acted beneficently is indeed a coherent possibility, acting on a
maxim of non-beneficence does not violate WU. On Kant’s view, of course, acting on
such a maxim does run afoul of the Formula of Universal Law. It does so, he thinks,
because as a rational agent it is not possible to act on it and, at the same time, will that its
generalization be a universal law. On the basis of the criteria discussed thus far, Kant
does not appear to have the tools to eliminate WU as a contender for the supreme
principle of morality. For not only is it possible that WU satisfies Kant's basic concept of
the supreme principle of morality, but there seems to be no reason to think that it couldn’t
fulfill the criteria suggested by his three propositions.

In my view, this difficulty prompts us to see that Kant actually suggests one
further criterion for the supreme principle of morality. It must be such that a plausible set
of duties, that is, plausible relative to common rational knowledge of morality, would
stem from the principle. Both BP and WU could be eliminated through an appeal to this
criterion. According to ordinary moral thinking, contrary to BP and to WU, we have a
duty of beneficence.

A textual basis for this criterion is not hard to discern. In Groundwork II, Kant
offers a derivation of the Formula of Universal law that parallels his derivation in
Groundwork I. Right after he arrives at this formula, Kant says: “Now, if all imperatives
of duty can be derived from this one imperative as their principle, then, even though we
leave it undecided whether what is called duty is not as such an empty concept, we shall
still be able to indicate at least what we understand by it and what the concept means”
(GMS 420-421). The derivation is not complete unless “all imperatives of duty” can be
derived from the imperative Kant proposes as the only viable candidate for the supreme
principle of morality. By “all imperatives of duty,” Kant apparently means all imperatives that we, reflective rational agents, take to express our moral duties. Kant proceeds, of course, to try to show that four such imperatives, including, for example, a requirement not to make false promises for financial gain, follow from the Formula of Universal Law. He then says: “These are some of the many actual duties—*or at least of what we take to be actual*—whose derivation from the single principle cited above is perspicuous” (GMS 424, italics mine). If these duties’ derivation from the Formula of Universal Law were not clear, for example, if it simply did not follow from the formula that we had them, then, Kant implies, we could not accept this formula as the only viable candidate for the supreme principle of morality. In the short paragraph (GMS 420-421) following his statement of the Formula of Universal Law, Kant indicates an important criterion for any viable candidate for the supreme principle of morality. We must be able to see how it follows from this candidate that, if it were established, we would indeed have moral duties that we are convinced we do have. (For further textual evidence that Kant embraces this criterion, see Kerstein 2002: 87-89.)

With this additional criterion in place, Kant can advance towards eliminating rivals for status as viable candidates for the supreme principle of morality. But in order for his derivation of the Formula of Universal Law to succeed, he would need not only to show that rivals are unfit to satisfy the criteria he indicates, but also that it remains viable to think that his candidate can satisfy them. Yet this latter task poses serious challenges with respect to the additional criterion. Kant offers various, supposedly equivalent, formulations of the categorical imperative. For example, in addition to the Formula of Universal Law, he offers the Formula of Humanity: “Act in such a way that you treat
humanity, whether in your own person or in any other person, always at the same time as an end, never merely as a means” (GMS 429, emphasis omitted). Both of these principles are difficult to interpret; it’s far from obvious what either one would require us to do. In my view, it is very unlikely that the Formula of Universal Law would generate a set of duties acceptable to ordinary moral thinking (Kerstein 2002: 168-174). The Formula of Humanity seems more promising on this score, but on interpretations suggested recently by Kantians it too has implications that fail to square with common notions of morality (Kerstein, 2002: 177-187).

In *Groundwork* I, Kant offers a derivation of the Formula of Universal Law. He tries to show that if there is a supreme principle of morality, then it is this principle (or its equivalent.) He suggests a three step process towards attaining this goal, I believe. The first step is to develop criteria for the supreme principle of morality. In order to do this, Kant appeals at key points to common rational knowledge of morality. He bases his notion that the supreme principle must have “absolute necessity” on such an appeal. The criterion implicit in his first proposition, as well as the criteria implied in his second and third propositions, stem ultimately from his notion that, according to ordinary moral thinking, an action has moral worth if and only if it is done from duty. The last criterion we discussed appeals directly to the moral verdicts of common sense. The Formula of Universal Law is, of course, not the only principle that philosophers have sought to elevate to status as the supreme principle of morality; Kant’s principle has many rivals. The second step in the derivation is to eliminate these rivals on the basis of their manifest inability to fulfill all of the criteria. Finally, Kant has to show that the Formula of
Universal Law remains as a viable candidate for fulfilling all of the criteria. (He does not have to prove that the Formula of Universal Law actually does fulfill the whole set. For that would entail establishing that this principle is absolutely necessary and universally valid. And that is a project that Kant undertakes in the notoriously difficult third chapter of the *Groundwork*.) Each one of these steps is controversial. But, in my view, Kant nevertheless offers a coherent, philosophically interesting argument for his conclusion that either there is no supreme principle of morality or it is the Formula of Universal Law.

VI

Let me close by considering an objection to this account of Kant’s derivation. The objection is that it is not consistent with Kant's claim the supreme principle of morality must be an a priori principle. In particular, consider the criterion according to which the supreme principle must be capable of generating duties that cohere with the moral duties we take ourselves to have. Does not whether we conclude that a given principle meets this criterion rest on experience, that is, our particular experience of morality? Already in the *Groundwork* Preface Kant says that the ground of an obligation to conform to the supreme principle of morality must be sought “solely a priori in concepts of pure reason” and that any principle “can indeed be called a practical rule, but never a moral law, so far as it rests even slightly (perhaps only in its motive) on empirical grounds” (GMS 389).

In order to respond to this objection, we need to understand two senses in which, according to Kant, the supreme principle must be an a priori rather than an empirical
principle. It must be a priori in both (what I call) a motivational sense and an epistemological sense.\(^3\)

Beginning with the former, the supreme principle of morality must be such that all rational agents always have available to them a sufficient motive for abiding by it. (Whether they actually act on this motive or some other one, such as an inclination, is another question.) But that means that their having sufficient motive available to them to conform to the principle must not depend on anything empirical, that is, on their particular inclinations or even on their nature, insofar as this nature is not necessarily shared with all rational agents. A principle is a priori in the motivational sense just in case any rational agent’s having available to him a sufficient motive for abiding by it is not conditional on anything empirical. (In effect, a principle is a priori just in case it fulfills the criterion implicit in Kant’s third proposition). A practical principle would be empirical, for example, in case a rational agent’s having sufficient motive to abide by it was conditional on his expectation that abiding by it would give him pleasure (KpV 9, note).

Kant’s appealing to experience in his derivation of the Formula of Universal Law does not seem incompatible with all rational agents having an empirically unconditioned motive at their disposal for abiding by this formula. That we rely on our moral experience in pinpointing the supreme principle of morality does not, for example, seem to entail that our having at our disposal sufficient motive to comply with it is conditional on our expectation that doing so will get us something we want.

\(^3\) For a different account of Kant’s emphasis on the a priori in the development of his moral philosophy, see Hill, 2002: 15-22.
The second sense in which, according to Kant, the supreme principle of morality must be a priori is what I call the epistemological sense. Kant states that a practical law, and thus the supreme principle of morality, must be knowable a priori (KpV 26; see also GMS 440). In the *Critique of Pure Reason*, Kant defines a priori knowledge as “knowledge absolutely independent of all experience” (KrV B 3). If we had a priori knowledge of a judgment or proposition, this knowledge would have to be “absolutely independent” of all experience in the following sense: it would have to be *grounded or legitimated* without appeal to any particular set of experiences (see Allison 1983: 78). So, it seems, if we had a priori knowledge of the supreme principle of morality, that is, if we knew that it was necessarily binding on all rational agents, this knowledge would likewise have to be grounded or legitimated without appeal to any particular set of experiences (see GMS 389).

Does the account sketched above of Kant’s *Groundwork* I derivation clash with the notion that the supreme principle of morality must be a priori in this sense? I do not believe so. But before explaining why, let me make a preliminary point. It would be a mistake to maintain that in *Groundwork* I Kant proves entirely a priori (or could reasonably take himself to prove entirely a priori) that if there is a supreme principle of morality, then it is the Formula of Universal Law. Maintaining this would be a mistake even for those who reject the notion that, according to Kant, a criterion for the supreme principle is that it generate duties in line with those that we take ourselves to have.

Kant’s first proposition is a cornerstone of the derivation on any plausible interpretation. But in my view Kant neither establishes the first proposition a priori nor could he reasonably take himself to do so. According to this proposition, an action has
moral worth if and only if it is done from duty. If the proposition were analytic, that is, if
the notion of an action’s having moral worth contained the notion of its being from duty
as the notion of something’s being a body contains the notion of its being extended, then
it would not be necessary to appeal to any particular experiences in order to justify it.
Kant would need only to analyze the concept of moral worth in order to demonstrate that
all and only actions done from duty have such worth. This analysis would reveal it to be
self-contradictory to claim that actions done from sympathy have moral worth. I do not
find such analysis in *Groundwork* I. (Why, precisely, would it be incoherent to think of
an action done from sympathy as unconditionally and preeminently valuable?) Although
it might be mistaken to claim that actions from sympathy have moral worth, it is far-
etched to maintain this claim to be self-contradictory.

Since it is implausible to maintain that the first proposition is analytic, it would be
no simple matter to justify it without appeal to any particular experiences. So far as I can
tell, Kant’s justification of the first proposition does appeal to particular experiences. To
cite one example, in his argument for the claim that only actions done from duty have
moral worth, Kant relies on the idea that any motive besides duty sometimes generates
actions contrary to duty (see II above). But, at least implicitly, he seems to justify this
idea through an appeal to particular experiences, namely experiences of motives other
than duty actually producing actions contrary to duty (or at least contrary to what we
believe our duty to be). \(^4\) Since Kant does not ground his first proposition a priori, and this

\(^4\) One might dispute my contention that in his argument for the claim that only actions
done from duty have moral worth, Kant appeals to particular experiences, for example,
our experience that motives other than duty sometimes generate actions contrary to duty
proposition is key to his *Groundwork* I derivation of the Formula of Universal Law, he
does not there offer a purely a priori justification of his contention that if there is a
supreme principle of morality, then it is the Formula of Universal Law.

That is not surprising; for Kant does not assert that his *Groundwork* I derivation
rests solely on a priori grounds.\(^5\)\(^6\) He does, of course, claim that the supreme principle of

---

\(^5\) It might seem that Kant asserts this at the very beginning of *Groundwork* II. “Although
we have drawn our previous concept of duty form the common use of our practical
reason,” he says, “this by no means implies that we have treated it as a concept derived
from experience” (GMS 406). But Kant is not here asserting that the *Groundwork* I
derivation rests entirely on a priori grounds. In the discussion that follows, Kant suggests
that to treat the concept of duty as a concept of experience is to claim that it stems solely
from experience. But, he argues, the concept of duty (or, more precisely, the concept of
acting from duty) as described in *Groundwork* I cannot stem solely from experience. For,
according to this concept, when an agent acts from duty he does something solely
because it is required by moral principle. But experience is unable to show us even a
single case in which an agent does something solely on these grounds. According to
Kant, the concept of acting from duty implicit in ordinary moral thinking does not derive
simply from experience. Of course, the claim, defended above, that the *Groundwork* I
morality must be knowable a priori. It must be possible to have a certain kind of knowledge that it is valid, namely knowledge that is not based on appeals to particular experiences. The *Groundwork* I derivation’s being based partly on appeals to experience does not itself undermine this claim. That in *Groundwork* I we reflect on our experience derivation rests *partly* on appeals to experience does not entail that this derivation is based *exclusively* on such appeals. The interpretation offered here is consistent with Kant’s view that the concept of acting from duty implicit in ordinary moral thinking does not derive simply from experience.

Does Kant believe that his *Groundwork* II derivation of the Formula of Universal Law (which culminates around GMS 421) is a purely a priori argument, that is, one that depends not at all on appeals to particular experiences? An adequate answer to this question would require far more discussion than is appropriate here. But let me offer a brief response. Kant does sometimes suggest that he is committed to the *Groundwork* II derivation being purely a priori (see, e.g., GMS 412). Yet it is questionable whether this is his considered view. He implies that the derivation is not complete until he shows that the Formula of Universal Law generates moral duties we take ourselves to have (GMS 421). And showing this involves an appeal to experience, namely to the fact that we hold various sorts of action (e.g., false promising solely for financial gain) to be morally impermissible. If we do interpret Kant’s *Groundwork* II derivation as a purely a priori argument, I fear that we must conclude it to be a miserable failure. For I do not see how, without appealing to an empirically-grounded criterion for the supreme principle of morality such as the one introduced in V above, Kant is able to show that a whole range of competitors to the Formula of Universal Law are unsuited for status as this principle.
in order to see that a certain principle is at work in our moral practice does not entail that we cannot know this principle a priori.

In *Groundwork* III, Kant attempts to establish the Formula of Universal Law (or at least a principle closely resembling it). He tries to show that it would be irrational for any being within its scope, that is, any rational agent, to fail to comply with it. Any argument that proved this, Kant believes, could not be based on appeals to experience. The argument of *Groundwork* III is difficult to pinpoint. But if we assume that the argument does not (even indirectly) appeal to particular experiences, then the following point becomes evident. In establishing the Formula of Universal Law, Kant would, in effect, show that it is knowable a priori. For he would show something stronger, namely that it is known a priori, at least by those who understand the argument.
References


