

FROM THE GOOD WILL TO THE FORMULA OF UNIVERSAL LAW¹

Samuel C. Rickless

1. Introduction

In the First Section of the *Groundwork of the Metaphysics of Morals*, Kant argues that a good-willed person “under subjective limitations and hindrances” (G 397) is required “never to act except in such a way that [she] could also will that [her] maxim should become a universal law” (G 402).² This requirement has come to be known as the Formula of Universal Law (FUL) version of the Categorical Imperative, an “ought” statement expressing a command of reason that “represent[s] an action as objectively necessary of itself, without reference to another end” (G 414). The question of how to understand and apply the FUL has received a great deal of attention in recent years. But Kant’s *reasoning* for the claim that a good-willed person (of a certain kind) acts in accordance with the FUL has not been investigated with the same degree of thoroughness. My purpose in this paper is to render the structure of Kant’s argument perspicuous, and thereby contribute to its proper evaluation by identifying its true strengths and weaknesses.³

Let us call Kant’s argument “K” and its conclusion “C”. The task at hand requires identification of K’s premises and of the reasoning by means of which C is derived. One of the advantages of clarifying K’s structure in this way is that doing so permits us to answer questions that continue to puzzle even the most sympathetic readers of the First Section. Some of these questions concern three “Propositions” to which Kant

draws our attention in the course of his presentation. The First Proposition (call it “P1”) is left unstated, but most commentators agree that it (or at least part of it) should be rendered as follows:

(P1) A human action has moral worth only if it is done from duty.⁴

Kant states the other two propositions (call the Second “P2” and Third “P3”) explicitly:

(P2) An action from duty has its moral worth *not in the purpose* to be attained by it but in the maxim in accordance with which it is decided upon.

(G 399)

(P3) Duty is the necessity of an action from respect for law. (G 400)

The first question occasioned by Kant’s having drawn our attention to these Propositions is what role he takes them to play in K. That Kant does not consider the Third to be one of K’s premises is clear from his remark that P3 “is a consequence of” P1 and P2 (G 400). But does Kant take P1 and P2 to be (unargued for) premises of K, or does he take them to be conclusions based on more fundamental statements asserted earlier in the First Section?

Secondly, Kant’s claim that P3 follows from the conjunction of P1 and P2 has been met with some scepticism, no more forcefully articulated than by Wood (1999, 43):

Kant says that [P3] “is a consequence of the two preceding,” but he cannot mean that it is logically entailed by them, for it plainly is not. In fact, at the beginning of the next paragraph he seems to be inferring [P2] from [P3] (G 4:401). More likely he means merely that [P3] is a consistent extension of [P1] and [P2] or lies a bit farther along the same path of thinking down which he has been directing us with [P1] and [P2].

So our next question is whether P3 is, as Kant suggests, logically entailed by P1 or P2 (perhaps in conjunction with other assumptions that Kant has been taking for granted).

Third, what is, and what does Kant believe to be, the logical relationship (if any) that obtains among P1, P2, P3, and C? Do P1 and P2 play any role in the derivation of C? And does Kant believe (and is it in fact true) that C follows from P3 (perhaps in conjunction with other assumptions)?

Connected to this question is the problem of the logical “gap” in the derivation of FUL (at G 402) to which Aune (1979) first drew attention. Although several commentators have attempted to “fill” the gap with reasoning which, they theorize, Kant accepted (or would have accepted if it had been brought to his attention), no consensus has yet been reached on the question of whether any of these attempts succeeds.⁵ We are then left to wonder whether it is at all possible to construct a persuasive line of reasoning from the good will to FUL that Kant would (or at least, might) have accepted.

The purpose of this paper is to provide a precise reconstruction of K that will enable us to answer the questions just raised. My plan is to proceed as follows. In section 2, I consider whether Kant treats P1 as an independent premise or as the

conclusion of an argument based on more fundamental considerations. I argue that P1 is the conclusion of an argument, and identify the fundamental premises from which Kant attempts to derive it. In section 3, I reconstruct Kant's argument for P2, and in section 4, Kant's argument for P3. In section 5, I describe the "gap" in Kant's argument from P3 to C. In section 6, I propose a simple, plausible argument for C that relies on statements made in the Second Section and in the *Critique of Pure Reason*. Section 7 then concludes with a restatement of the main lines of K and a summary of results obtained from the investigations conducted in all preceding sections.

2. From the Good Will to the First Proposition

To understand why Kant takes P1 to be true, it helps to begin with the work of Potter 1974. Potter identifies two propositions asserted by Kant early in the First Section. The first (call it "T1") is stated at the very outset (at G 393): "It is impossible to think of anything at all in the world, or indeed beyond it, that could be considered good without limitation except a *good will*." Potter restates T1 as follows:

(T1) A good will only [i.e., is the only thing that] has absolute worth.

What allows Potter to use the phrase "absolute worth" to characterize a good will Kant himself describes as "good without limitation" at G 393 is the fact that Kant uses (the German equivalent of) such terms interchangeably throughout the *Groundwork*. At G 393-394, for instance, Kant characterizes "moderation in affects and passions, self-control, and calm reflection" as having no "unconditional worth" (*unbedingten Werth*)

and equates this both with their failing to be “absolutely good” (*schlechthin gut*) and with their failing to be “good without limitation” (*ohne Einschränkung für gut*). He also describes the good will as having “absolute worth” (*absoluten Werthe*) at G 394, strongly suggesting that he identifies the concepts of absolute goodness and absolute worth.⁶

The fundamental idea here, as revealed in Kant’s own illustrations of T1 at G393, is that anything other than the good will is conditionally good, in the sense that its goodness is conditional (depends) upon the goodness of something else, whereas the good will is itself unconditionally good, in the sense that its goodness is not conditional (does not depend) upon the goodness of anything else. Kant then relies on the reader’s ability to recognize that the condition on which the goodness of something depends may reasonably be described as “limiting” that goodness and as something to which that goodness is relative, and hence not absolute. And this is why Kant describes the unconditionally good will as “good without limitation” and “absolutely good”.

The second thesis (call it “T2”) is stated at the start of the third paragraph (at G 394):

(T2) A good will is not good because of what it effects or accomplishes.

Potter goes on to say that “Kant gives little in the way of argument for [T1]” (30), that T2 follows from T1 (31), and that “Kant thought that [P1] followed from [T1], though the way in which it follows is not obvious” (32). Moreover, claims Potter, the unobvious “transition” from T1 to P1 is in two moments: first, a “shift from the value of a good will, to the value of actions which are expressions of that will”, and second, “the introduction

of the concept of duty, as a concept central to the analysis of acts expressive of a good will in human beings” (32).

Much of what Potter says here is, in my view, absolutely correct. First, Potter is right to claim that T2 follows from T1. For to say that X is unconditionally good is to say that X does not depend for its goodness on the goodness of anything else. But if X were good “because of what it effects or accomplishes”, then X would depend for its goodness on the goodness of its effects. Therefore, if a good will is unconditionally good, then it isn’t good because of what it effects or accomplishes. Second, as we shall see, Potter rightly points out that the move to P1 does depend on the two moments identified in the previous paragraph.

But there are two problems with Potter’s reconstruction of Kant’s argument for P1. In the first place, some of what Kant says can be put together in the form of an argument for T1. So it is somewhat misleading to say that Kant gives “little in the way of argument” for T1. And second, Kant’s argument for P1 rests not on T1, but on the fundamental premise from which T1 can be derived. Let us then consider each of these claims in turn.

It might seem that Kant’s stating T1 at the very outset of the First Section indicates that he treats it as the most fundamental assumption of the *Groundwork*. But since it is not rare for an author to present his conclusion at the outset and then proceed to argue for it, we must look at the rest of the Section to see whether Kant provides any support for T1. Upon further reflection, the following statements leap to the eye:

A good will seems to constitute the indispensable condition even of the worthiness to be happy. (G 393)

[The good will] must still be the highest good and the condition of every other, even of all demands for happiness. (G 396)

[The concept of a good will] always takes first place in estimating the total worth of our actions and constitutes the condition of all the rest. (G 397)

Kant is making two claims here. The first claim is this:

(1) A good will is the condition of every other good.

I take (1) to mean that every good thing other than a good will depends for its goodness on its being associated with a good will. The second is a particular instance of the first, namely that a person's worthiness to be happy depends for its goodness on her having a good will. And this is not the only instance of (1) to which Kant draws our attention. In the first two paragraphs of the First Section, he tells us that talents of mind (such as understanding and judgment), qualities of temperament (such as courage and moderation), and other gifts of fortune (such as power, wealth, and health) are good only when they are associated with a good will. This fact explains our finding the coolness of a scoundrel "abominable" (G 394) and our taking "no delight in seeing the uninterrupted prosperity" of a bad-willed person (G 393).

None of this is new or surprising. But one of the more surprising facts worth noting is that (T1) follows from (1) (when conjoined with one implicit assumption). The argument is as follows. (1) says that every good thing other than a good will depends for its goodness on a good will. Assuming, as seems reasonable, that there must be some good thing that does not depend for its goodness on the goodness of anything else (for otherwise we would be forced into circularity or infinite regress), it follows directly that this thing must be a good will. (Were it anything other than a good will, (1) tells us that it would depend for its goodness on something else.) Thus, a good will is unconditionally good, and it is the only thing that is unconditionally good. So (1), along with the claim that there must be some good thing that does not depend for its goodness on the goodness of anything else, entails T1.

It is therefore more appropriate than not to read the First Section of the *Groundwork* as containing an argument for T1. Even more interesting, in my view, is that Kant's argument for P1 relies not on T1, but on (1). (1) says that every good thing other than a good will depends for its goodness on its being associated with a good will. It follows directly that all good (morally worthy) human actions depend for their goodness (moral worth) on the good will with which they are associated (i.e., the good will that performs them), and therefore that human actions have moral worth only if they are done by a good will. Call this result "(2)":

(2) A human action has moral worth only if it is done by a good will.

This is the first moment in the argument to P1 identified by Potter, namely the “shift” from the value of the good will to the value of the actions done thereby. Now for the second, and more familiar, moment: the introduction of the concept of duty. Kant claims that “the concept of duty...contains that of a good will though under certain subjective limitations and hindrances” (G 397). From this “concept containment” thesis, Kant takes it to be (analytically) the case that any action performed by someone who possesses inclinations (those bothersome “limitations and hindrances”) is done by a good will if and only if it is done from duty. Since humans possess inclinations, the thesis tells us that

(3) A human action is done by a good will if and only if it is done from duty.

P1, the claim that a human action has moral worth only if it is done from duty, then follows directly from (2) and (3).

Here, then, is the logical lay of the land. Kant had the resources to infer T1 from (1), and then inferred T2 from T1. But with respect to Kant’s argument for P1, T1 and T2 are both otiose. For Kant uses (1) (not T1 or T2) to derive (2), asserts the analytic truth of (3), and then infers P1 from the conjunction of (2) and (3). So the soundness of Kant’s argument for P1 depends primarily on the truth of (1) and (3).⁷

Having established P1, Kant goes on to provide various now famous examples of morally required actions divided into three categories: actions done from indirect inclination (i.e., self-interested actions), actions done from direct inclination (e.g., helpful actions done out of sympathy for those in need), and actions done from duty. There is a

significant amount of controversy concerning the proper interpretation of these examples.

On the reading I favor, Kant uses the examples to make two points:

(A1) Actions done from (direct or indirect) inclination have no moral worth.

(A2) Actions done from duty have moral worth.

But (A1) and (A2) together entail that actions cannot be performed from duty and from inclination at the same time, and some scholars find it uncharitable to foist such an unreasonable psychological thesis on Kant without further textual or theoretical support. To bolster their case, these scholars may suggest that Kant's examples be read as supporting, not (A1) and (A2), but (A1)* and (A2)*:

(A1)* Actions done *from inclination alone* have no moral worth.

(A2)* Actions done *from duty alone* have moral worth.

And they may go on to point out that (A1)* and (A2)* simply do not speak to a possibility that Kant, for all we know, took very seriously, namely the case in which an action is overdetermined in the sense that it is done both from inclination and from duty.⁸

A thorough discussion of this issue is beyond the scope of this paper. But there is significant textual evidence, both direct and indirect, that Kant (at least in the *Groundwork*) did not allow for the possibility of overdetermination. First, Kant tells us explicitly that “an action from duty is to put aside entirely the influence of inclination” (G 400), and this strongly suggests that Kant took an action's being done from duty to entail

its not being done from inclination.⁹ Second, as I will argue in the next two sections, it becomes necessary to read Kant as excluding the possibility of overdetermination (in the relevant sense) if one wishes to read his argument for P3 as valid. On my view, these points, taken together, indicate that the available evidence weighs more heavily on the side of the claim that Kant intended his now famous examples to establish (A1) and (A2).¹⁰

3. From the First Proposition to the Second Proposition

Having established and illustrated P1, Kant immediately supplies us with an argument for P2 contained in the following passage (at G 400), which I have divided into two parts, (α) and (β):

- (α) The second proposition is this: an action from duty has its moral worth *not in the purpose* to be attained by it but in the maxim in accordance with which it is decided upon, and therefore does not depend upon the realization of the object of the action but merely upon the *principle of volition* in accordance with which the action is done without regard for any object of the faculty of desire. That the purposes we may have for our actions, and their effects as ends and incentives of the will, can give actions no unconditional and moral worth is clear from what has gone before. In what, then, can this worth lie, if it is not to be in the will in relation to the hoped for effect of the action? It can lie nowhere else *than*

in the principle of the will without regard for the ends that can be brought about by such an action.

- (β) For, the will stands between its a priori principle, which is formal, and its a posteriori incentive, which is material, as at a crossroads; and since it must still be determined by something, it must be determined by the formal principle of volition as such when an action is done from duty, where every material principle has been withdrawn from it.

Reacting to these admittedly dense passages, Potter (33-34) claims that “Kant is not entirely clear about what he is arguing for,” inasmuch as “he is establishing two distinct propositions” (call them “P2a” and “P2b”), the first being the conclusion of (α), the second the conclusion of (β):

(P2a) The moral value of an action is a quality of its maxim.

(P2b) Action from duty is action done on the basis of a maxim which we have adopted, not because its adoption will fulfil desires of ours, but because of its form.

Potter then claims that P2b is an essential premise of K, but that P2a plays no further role in Kant’s reasoning.

These claims, I will now argue, are right in certain respects, but wrong in others. On the one hand, it is true that *something like* P2b is the conclusion of (β), that *something like* P2a is *part of* the conclusion of (α), that the conclusion of (β) is a crucial step in K, and that the conclusion of (α) is not. On the other hand, I think it does Kant a disservice to say that he does not make clear what he is arguing for. Upon close analysis, it becomes evident that the arguments in (α) and (β) are not logically independent, for the conclusion of (β) is an important premise in the argument for the conclusion of (α). What Potter fails to realize is that, taken together, (α) and (β) contain, not two arguments for two unconnected conclusions, but rather a single sustained argument for P2.

In order to reveal Kant's reasoning in support of P2, let us begin with the argument contained in (β). The conclusion of (β) (call it "C β ") is that a human will that performs an action from duty is determined to perform the action, not by an *a posteriori* incentive, but by an *a priori* principle. (Notice that C β is *something like* Potter's P2b.) The main lines of the argument for C β are these. Kant tells us, first, that the will "stands between its *a priori* principle...and its *a posteriori* incentive...as at a crossroads." By this, I take Kant to mean that the human will is determined to perform its actions either by an *a priori* principle or by an *a posteriori* incentive. Kant then claims that "every material principle has been withdrawn from" a human will when it performs an action from duty. By this, I take him to mean that a human will that performs an action from duty is not determined to perform the action by any material principle. He then concludes that a human will that performs an action from duty is determined to perform it, not by an *a posteriori* incentive, but by an *a priori* principle.

The first thing to notice about this argument outline is that the conclusion does not follow if we are not entitled to conclude from the fact that the will is not determined by any material principle that it is not determined by an *a posteriori* incentive. Now Kant does tell us in (β) that *a posteriori* incentives are material. So we would be able to conclude that W is not determined by an *a posteriori* incentive, if we were entitled to assert that W is not determined by a material *incentive*. But what Kant actually *says* is that, in the case of human actions done from duty, the will is not determined by a material *principle*. Further, at G 427-28, Kant tells us that a *principle* is material when it has put a subjective end, and consequently a certain incentive, at its basis. (Kant says “consequently” because, on his view, subjective ends “rest on incentives” (G 427), and any principle that is based on X is *ipso facto* based on anything on which X rests.) Thus, in order to conclude that W is not determined by a material *incentive*, Kant must assume that a will’s being determined by a material *incentive* is just its being determined by a material *principle* that is ultimately based on the incentive.

But by what right does Kant say that a human will that performs an action from duty is not determined to perform it by a material *principle*? The answer comes in three stages. First, for a human will W to be determined to perform A by a material principle is just for W to perform A from a desire for the subjective end on which the principle is based. Second, in Kant’s parlance, desires for subjective ends are just inclinations. And third, by A1 and A2, if an action is done from duty, then it is not done from inclination.

Here, then, is the full rendition of the argument for Cβ:

∴ (4) If a human action is done from duty, then it is not done from inclination.

[From A1, A2]

- (5) Desires for subjective ends are inclinations.
- (6) A human will is determined to perform an action by a material principle if and only if it performs the action from a desire for the subjective end on which the principle is based.
- ∴ (7) A human will that performs an action from duty is not determined to perform the action by a material principle. [From 4, 5, 6]
- (8) A human will is determined to perform an action by a material incentive if and only if it is determined to perform the action by a material principle that is ultimately based on the incentive.
- (9) All *a posteriori* incentives are material.
- (10) A human will is determined to perform its actions either by an *a priori* principle or by an *a posteriori* incentive.
- ∴ (Cβ) A human will that performs an action from duty is determined to perform the action, not by an *a posteriori* incentive, but by an *a priori* principle.
[From 7, 8, 9, 10]

That Cβ is itself a premise in an argument for a further conclusion is evidenced by the fact that (β) begins with the paradigmatic premise-indicator, “for”. The conclusion that Kant uses Cβ to support is stated at the very end of (α), namely that the moral worth of a human action done from duty “lies nowhere else than in the principle of the will without regard for the ends that can be brought about by such an action.” This claim, the conclusion of (α) (call it “Cα”), is just a stylistic variant of P2, for Kant uses the term

“maxim” in this context to mean the same as “principle of the will without regard for the ends that can be brought about by such an action.”¹¹ (Notice that *part of C α* is *something like* Potter’s P2a.) So, having established the truth of C β , Kant now wants to use C β to establish the truth of C α (=P2). But how?

The answer is this. First, from “what has gone before” it is clear that P1 is true, namely that a human action has moral worth only if it is done from duty. And second, Kant assumes that the moral worth of a morally worthy action depends on the motive for its performance. In other words, if an action has moral worth, then its moral worth lies in that which determines the will to perform the action. But, by C β , a human will that performs an action from duty is determined to perform the action, not by an *a posteriori* incentive, but by an *a priori* principle. It follows that the moral worth of a human action done from duty lies, not in an *a posteriori* incentive, but in an *a priori* principle. But this is just to say that the moral worth of a human action done from duty lies “nowhere else than in the principle of the will without regard for the ends that can be brought about by such an action.”

More formally, then, the argument for P2 is as follows:

(P1) A human action has moral worth only if it is done from duty.

(C β) A human will that performs an action from duty is determined to perform the action, not by an *a posteriori* incentive, but by an *a priori* principle.

(11*) If an action has moral worth, then its moral worth lies in that which determines the will to perform the action.¹²

∴ (12*) The moral worth of a human action done from duty lies, not in an a

posteriori incentive, but in an *a priori* principle. [From P1, C β , 11*]

∴ (P2) The moral worth of a human action done from duty lies nowhere else than in the principle of the will without regard for the ends that can be brought about by such an action. [=C α : Restatement of 12*]¹³

Here, then, is the logical lay of the land in the argument for P2. On the basis of a number of explicit and implicit premises (namely, A1, A2, 5, 6, 8, 9, and 10), Kant establishes the truth of C β . Kant then uses P1, C β , and an additional premise (namely, 11*) to establish the truth of C α (=P2). The upshot is that P2 is not an unargued-for, independent premise of K. Rather, Kant provides a valid argument for P2 that depends in large part on P1. All of this makes sense, for it helps to explain why Kant placed as much emphasis as he did on P1, and why he did so *before* undertaking his proof of P2.

4. The Argument for the Third Proposition

Having established P2, Kant now tells us (famously) that P3 “is a consequence of” P1 and P2. Is this anywhere near correct?¹⁴

P3 says that “duty is the necessity of an action from respect for law.” By this, I take Kant to mean the following: a (human) action is done from duty only if it is a morally required (“necessary”) action done from respect for law. Now Kant tells us (at G 401, fn.) that respect “signifies merely consciousness of the *subordination* of my will to a law without the mediation of other influences on my sense,” that it is consciousness of “immediate determination of the will by means of the law,” and is therefore “the *effect* of the law on the subject.” Thus, a will that does an action from respect for law is really just

a will that is (immediately and consciously) determined to perform the action by a law.

P3 may therefore be restated as follows:

(P3r) A human will performs an action from duty only if it is (immediately and consciously) determined to perform it by a law.

Kant's argument for P3r is contained in the following passage (at G 400-401), which I will call "γ":

(γ) Only what is connected with my will merely as ground and never as effect, what does not serve my inclination but outweighs it or at least excludes it altogether from calculations in making a choice—hence the mere law for itself—can be an object of respect and so a command. Now, an action from duty is to put aside entirely the influence of inclination and with it every object of the will; hence there is left for the will nothing that could determine it except objectively the *law* and subjectively *pure respect* for this practical law, and so the maxim of complying with such a law even if it infringes upon all my inclinations.

Start with the claim that "an action from duty is to put aside entirely the influence of inclination and with it every object of the will." Kant is here harking back to (4), namely that a human action is done from duty only if it is not done from inclination. But I take it that, in referring back to (4), Kant is also drawing our attention to what he ultimately used

(4) to prove, namely $C\beta$. Now $C\beta$ tells us that a human will performs an action from duty only if it is determined to perform the action by an *a priori* principle. To obtain P3r, all that Kant needs to do is to show that the *a priori* principle that determines a human will to perform actions that are done from duty is a *law*.

The argument for this claim is simple. In (γ), Kant tells us that a law is “what is connected with [one’s] will merely as ground and never as effect.” In (β), he says that *a priori* principles are “formal”, and later (at G 427-428) defines formal principles to be those principles that “abstract from all subjective ends.” But this is just to say that a principle is *a priori* only if it is never connected to the will as effect. Assuming that a principle of the will is connected with it either as ground or as effect, it follows that a principle is *a priori* only if it is connected with the will merely as ground and never as effect. But then the *a priori* principles that determine human wills to perform actions that are done from duty must be *laws*.

Here, then, is the argument for P3:

- ($C\beta$) A human will that performs an action from duty is determined to perform the action, not by an *a posteriori* incentive, but by an *a priori* principle.
- (11) A principle is *a priori* only if it is never connected to the will as effect.
- (12) A principle is connected with the will either as ground or as effect.
- \therefore (13) A principle is *a priori* only if it is connected to the will merely as ground and never as effect. [From 11, 12]
- (14) A principle is connected with the will merely as ground and never as effect only if it is a law.

- ∴ (P3r) A human will performs an action from duty only if it is determined to perform it by a law. [From Cβ, 13, 14]
- ∴ (P3) Duty is the necessity of an action from respect for law. [Restatement of P3r]¹⁵

We are now in a position to see that Kant's claim that P3 follows from P1 and P2 is, as Wood points out, incorrect. The claim is incorrect because P3 follows, not from P1 or P2, but from Cβ (along with other assumptions). Moreover, note that Cβ is logically connected to P2 (inasmuch as Cβ is an important lemma in the argument for P2), and that P2 is logically connected to P1 (inasmuch as P2 follows from P1, along with other assumptions). It is in this sense that Wood is right to suggest that P3 is a "consistent extension" of P1 and P2, and that P3 "lies a bit further along the same path of thinking down which [Kant] has been directing us" with P1 and P2. P3 is, in fact, the result of extending part of the logical reasoning upon which P2 is based, reasoning that itself relies on P1.

5. The Gap

Having shown that a human will performs an action from duty only if it is determined to perform it by a law, Kant is entitled to conclude, on the basis of (3), that a good human will is determined to perform its actions by a law. He then asks (at G 402): "But what kind of law can that be, the representation of which must determine the will, even without regard for the effect expected from it, in order for the will to be called good

absolutely and without limitation?” His answer, which has occasioned much controversy, appears in the following passage (call it “ δ ”):

- (δ) Since I have deprived the will of every impulse that could arise for it from obeying some law, nothing is left but the conformity of actions as such with universal law, which alone is to serve the will as its principle, that is, *I ought never to act except in such a way that I could also will that my maxim should become a universal law*. Here mere conformity to law as such, without having as its basis some law determined for certain actions, is what serves the will as its principle...

The purpose of (δ) is to argue, on the basis of P3, that the law that determines the actions of a good human will is expressed by FUL. It is therefore an argument for C. Scholars of Kant, including most notably Aune (1979), have noticed that Kant’s reasoning from P3 to C is invalid. This is the so-called problem of the “gap” in the derivation of FUL.

Kant does provide material that might be used to fill this gap. Earlier, at G 401, he had said that a will’s being determined by a law is just its acting on the “maxim of complying with such a law”. So, from the fact that a good human will is determined to perform its actions by a law, Kant infers that a good human will acts on the principle of complying with a law, or, as he puts it in (δ), the principle of conforming its actions with universal law.¹⁶ Kant now adds that a good human will’s acting on such a principle cannot be motivated by the thought that doing so will satisfy some “impulse”, since in

that case the will would be acting from inclination, and not from duty. Rather, a good human will must act on the principle of conforming its actions to a law *as such*.

More formally:

- (3) A human action is done by a good will if and only if it is done from duty.
- (P3r) A human will performs an action from duty only if it is determined to perform it by a law.
- (15*) A human will is determined to perform an action by a law if and only if it acts on the principle of conforming its actions to a law.
- ∴ (16*) A good human will acts on the principle of conforming its actions to a law. [From 3, P3r, and 15*]
- (17*) If a human will acts on the principle of conforming its actions to a law, then it acts on the principle in order to satisfy some impulse or it acts on the principle as such.
- (18*) If a human will acts on a principle in order to satisfy some impulse, then it acts from inclination.
- (4) If a human action is done from duty, then it is not done from inclination.
- ∴ (19*) A good human will acts on the principle of conforming its actions to a law as such. [From 3, 4, 16*, 17*, 18*]

But this attempt to fill the gap is not complete. As Aune (28-34) emphasizes, Kant then infers from (19*) without explanation (by means of a simple “that is”) that a good human will acts in accordance with FUL: *I ought never to act except in such a way*

that I could also will that my maxim should become a universal law. In order for the inference to FUL to go through, the principle of conforming one's actions to a law as such needs to be identified with the principle that requires action only on maxims that can be willed to become universal laws. But what could possibly justify such an identification?

6. How to Fill the Gap: A Kantian Proposal

In the Second Section of the *Groundwork*, at G 420-421, Kant argues that FUL may be extracted from “the mere concept of a categorical imperative.” I believe that this argument provides us with the beginnings of a simple and significantly more persuasive defense of C.

Recall that P3r says that a human will performs an action from duty only if it is determined to perform it by a law. But a will's being determined to perform an action by a law is just its taking the objective necessity of the action as its reason to perform it (see G 412-413). Moreover, to take the objective necessity of an action as one's reason to perform it is to act on a principle that represents the action as objectively necessary. Since, as applied to imperfectly rational human wills, a principle of this sort is what Kant calls a “categorical imperative” (G 414), it follows that a human will performs an action from duty only if it acts on a categorical imperative.

Kant then asks “whether the mere concept of a categorical imperative may not also provide its formula containing the proposition which alone can be a categorical imperative” (G 420). The case for his affirmative answer to this question appears at G 420-421 (call the passage “ε”):

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

Kant has already established that a human will acts from duty just when it acts on a categorical imperative. The reasoning in (ε) establishing that FUL is the formula of this categorical imperative parallels the argument in (δ) that a good human will (i.e., one that acts from duty) acts on FUL. On the basis of P3r, (15*), (17*), (18*), (3), and (4) (see above, pp. 21-22), Kant argues that the categorical imperative requires us to conform our actions to a law as such. But since, as (ε) tells us, a law “contains no condition to which it would be limited,” Kant infers that conformity to a law amounts to conformity to the law’s universality. Kant then adds that conformity to universality is precisely what FUL requires of one’s maxims, and concludes that FUL must be the formula of the categorical imperative (and hence the principle on which a good human will acts).¹⁷

Unfortunately, the reasoning in (ε) is no less “gappy” than the reasoning in (δ). For Kant does not explain why conformity to a law amounts to conformity to the law’s universality, and, more importantly, does not explain why a maxim’s conformity to

universality is exactly what FUL requires. To say that conformity to universality is what FUL requires is to say that a maxim conforms to universality just when it is possible to *will* that the maxim be a universal law. But why not think that a maxim conforms to universality just when it is possible to *conceive* the maxim as a universal law? Kant does not say, and his failure to provide an answer to this question opens up yet another gap in his argument for FUL.

However, Kant's defeat contains the seeds of hope. For it is arguable that the concept of a categorical imperative does indeed contain FUL, though perhaps not exactly in the way Kant thought it did. An imperative, as Kant defines it, is "expressed by an ought" (G 413), and an "ought"-statement is categorical just when it is unconditional. Kant contrasts categorical imperatives with conditional "ought"-statements that "represent the practical necessity of a possible action *as a means* to achieving something else that one wills (or that it is at least possible for one to will)" (G 414), i.e., with hypothetical imperatives. Now whereas a hypothetical imperative, being conditional, applies only to those rational beings who will the end to which the imperative represents an action as means, a categorical imperative, being unconditional, necessarily applies to *all* imperfectly rational beings (see G 408 and 432). It is, in Kant's sense, an *objective* (or *objectively necessary*) principle (see G 413 and CPR 36). Thus, a categorical imperative tells me that I ought (unconditionally) to act in a certain way (i.e., on a certain kind of maxim), and its objectivity entails that any maxim I ought to act on is one that every imperfectly rational being ought to act on. In other words, a categorical imperative contains the following formula: "I ought to act only on those maxims that every imperfectly rational being ought to act on."

This formula is not FUL. But we can obtain FUL with the assistance of two plausible, arguably analytic, principles. The first is the famous “ought-implies-can” principle (call it “OC”), which Kant explicitly endorses in the *Critique of Pure Reason* (at A547-8/B575-6 and A807/B835):

(OC) If X ought to do A, then X can do A.

The second is a principle I think Kant would have accepted, though (as far as I know) he never states it explicitly (call it “OW”, for “ought-to-will”):

(OW) If every imperfectly rational being ought to do A, then (fictionally assuming that I know my will to be immediately effective) I ought to will that every imperfectly rational being do A.

I say that Kant would have accepted OW because I think OW would have struck him as obvious. For let us suppose that everyone ought to refrain from torturing children. Were I (*per impossibile*) to know that I am able to make it the case that everyone so refrains merely by willing it, how could it possibly be permissible for me not to will that everyone so refrain? If I didn't so much as will that no-one tortured children even though I knew I was in a position to protect children from such suffering by that very act of volition, wouldn't that be (and wouldn't Kant take it to be) obviously reprehensible?

With the help of OC and OW, the reasoning from the concept of a categorical imperative to FUL is straightforward. We've already shown that a categorical imperative

tells me to act only on those maxims that every imperfectly rational being ought to act on. But, by OW, the maxims that every imperfectly rational being ought to act on are maxims that I *ought to will* that every imperfectly rational being act on (on the fictional assumption that I know my will to be immediately effective). And, by OC, these maxims are such that (on the same fictional assumption) I *can will* that every imperfectly rational being act on them. But to say that every imperfectly rational being acts on a maxim is to say that the maxim has become a universal law (among imperfectly rational beings). Assuming that OC and OW are analytic, it follows that a categorical imperative contains the formula: “I ought to act only on those maxims that I (on the fictional assumption that I know my will to be immediately effective) can will to become universal laws.” And this formula is FUL.¹⁸

Here, then, is a more formal version of the reasoning from P3r to C that I am proposing on Kant’s behalf:

- (P3r) A human will performs an action from duty only if it is determined to perform it by a law.
- (15) A will is determined to perform an action by a law if and only if it takes the objective necessity of the action as its reason to perform it.
- (16) A will takes the objective necessity of an action as its reason to perform it if and only if it acts on a principle that represents the action as objectively necessary.
- (17) Any principle that represents an action as objectively necessary is a categorical imperative.

- ∴ (18) A human will performs an action from duty only if it acts on a categorical imperative. [From P3r, 15, 16, 17]
- (19) A categorical imperative says that I ought to act on a certain kind of maxim.
- (20) A categorical imperative is objective, i.e., applies to all imperfectly rational beings.
- ∴ (21) A categorical imperative says that I ought to act only on those maxims that every imperfectly rational being ought to act on. [From 19, 20]
- (OW) If every imperfectly rational being ought to do A, then (fictionally assuming that I know my will to be immediately effective) I ought to will that every imperfectly rational being do A.
- ∴ (22) A categorical imperative says that I ought to act only on those maxims such that I (fictionally assuming that I know my will to be immediately effective) *ought to* will that every imperfect rational being act on them. [From 21, OW]¹⁹
- (OC) If X ought to do A, then X can do A.
- ∴ (23) A categorical imperative says that I ought to act only on those maxims such that I (fictionally assuming that I know my will to be immediately effective) *can* will that every imperfect rational being act on them. [From 22, OC]²⁰
- (24) For every imperfectly rational being to act on a maxim is for the maxim to be a universal law.

- ∴ (25) A categorical imperative says that I ought to act only on those maxims that I (fictionally assuming that I know my will to be immediately effective) can will to be universal laws. [From 23, 24]
- ∴ (26) A human will performs an action from duty only if it acts on the principle: Act only on those maxims that you (fictionally assuming that you know your will to be immediately effective) can will to be universal laws. [From 18, 25]
- (3) A human action is done by a good will if and only if it is done from duty.
- ∴ (C) A good human will acts on the principle: Act only on those maxims that you (fictionally assuming that you know your will to be immediately effective) can will to be universal laws. [From 26, 3]

Notice that OW is a crucial element in this argument for C. For if we do not accept OW, then we must admit that morality requires much less of us than Kant (or common sense) suggests. Recall that Kant was able to show that a categorical imperative tells us to act only on those maxims that every imperfectly rational being ought to act on. If we eschew appeal to OW, then the most we can obtain from OC is the claim that every imperfectly rational being can act on maxims that every imperfectly rational being ought to act on. And from (24) we then obtain the following formula: “I ought to act only on those maxims that can become universal laws.” What this formula tells me to do is to restrict my maxims to those that pass what has come to be known as the “contradiction-in-conception test” portion of FUL. But this formula generates only one part of common-sense morality, and leaves out maxims that pass the contradiction-in-conception test but

fail what has come to be known as the “contradiction-in-the-will test.” Maxims that fail the latter test include, most prominently, maxims of non-beneficence.²¹ So if we wish to capture the whole (or at least the lion’s share) of common-sense morality, and explain the immorality of maxims of non-beneficence in particular, we must help ourselves to OW on Kant’s behalf.

Here, then, is the upshot: by supplementing Kant’s own statements about categorical imperatives in the *Groundwork* with OC and OW, we have provided strong support for the claim that FUL is indeed the formula of the fundamental principle of morality and the basis of the metaphysics of morals.

7. Conclusion

Without reproducing K in its entirety, let me now summarize the main lines of what I take to be the best argument for C that can be gleaned from the First Section of the *Groundwork*. Kant begins with (1), the proposition that a good will is the condition of every other good. On the basis of (3), the (for Kant, analytic) proposition that a human action is done by a good will if and only if it is done from duty, Kant infers the First Proposition, P1: A human action has moral worth only if it is done from duty. On the basis of further assumptions (A1, A2, 5, 6, 8, 9, and 10), Kant then establishes the truth of C β , namely that a human will that performs an action from duty is determined to perform the action, not by an *a posteriori* incentive, but by an *a priori* principle. From C β (along with 11, 12, and 14), Kant infers the Third Proposition, P3, which we restated as P3r: A human will performs an action from duty only if it is determined to perform it by a law. And on the basis of P3r, OW, and OC (together with 3, 15, 16, 17, 19, 20, and 24), we are

able to help Kant construct a valid argument for C, the conclusion that a good human will acts only on those maxims that it can (assuming it can universalize them) will to be universal laws.

Although Kant does not use P1 to argue for P3, he does use P1 (along with C β and 11*) to argue for his Second Proposition, P2: The moral worth of a human action done from duty lies nowhere else than in the principle of the will without regard for the ends that can be brought about by such an action. But P2 does no further work for Kant, and is consequently otiose in K. It is therefore arguable that Kant should have given pride of place to C β , rather than to P1 and P2 as he did. For it is on C β that the argument for P3 turns, and P3 is the central premise in the rest of Kant's argument for C.

If we add to all of this the previous section's Kant-inspired way of filling the notorious gap in the argument from P3 to C, then Kant's voyage from the good will to the Formula of Universal Law turns out to be more than just an easily criticized historical curiosity: it is a bold, rigorous, and quite possibly sound attempt to identify the fundamental moral principle on which good human wills act.²²

REFERENCES

Allison, H. E. 1990. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.

Allison, H. E. 1991. "On the Presumed Gap in the Derivation of the Categorical Imperative." *Philosophical Topics* 19. Reprinted in Allison 1996, 143-54.

Allison, H. E. 1996. *Idealism and Freedom: Essays on Kant's Theoretical and Practical Philosophy*. Cambridge: Cambridge University Press.

Aune, B. 1979. *Kant's Theory of Morals*. Princeton: Princeton University Press.

Baron, M. 1995. *Kantian Ethics Almost Without Apology*. Ithaca: Cornell University Press.

Henson, R. 1979. "What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Action." *Philosophical Review* 88: 39-54.

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

- Jensen, H. 1989. "Kant on Overdetermination, Indirect Duties, and Moral Worth."
In *Proceedings: Sixth International Kant Congress*, eds. G. Funke and T. M. Seebohm (Washington, D.C.: Center for Advanced Research in Phenomenology: University Press of America), 161-170.
- Korsgaard, C. M. 1985. "Kant's Formula of Universal Law." *Pacific Philosophical Quarterly* 66. Reprinted in Korsgaard 1996a, 77-105.
- Korsgaard, C. M. 1989. "Kant's Analysis of Obligation: The Argument of *Groundwork* I." *Monist* 72. Reprinted in Korsgaard 1996a, 43-76.
- Korsgaard, C. M. 1996b. "From Duty and for the Sake of the Noble: Kant and Aristotle on Morally Good Action." In *Aristotle, Kant, and the Stoics: Rethinking Happiness and Duty*, eds. S. Engstrom and J. Whiting, 203-36. Cambridge: Cambridge University Press.
- O'Neill, O. 1985. "Consistency in Action." In *Universality and Morality: Essays on Ethical Universalizability*, eds. N. Potter and M. Timmons, 159-86. Dordrecht: Reidel. Reprinted in *Constructions of Reason: Explorations of Kant's Practical Philosophy*, 81-104. Cambridge: Cambridge University Press, 1989.

Potter, N. 1974. "The Argument of Kant's *Groundwork*, Chapter 1." *Canadian Journal of Philosophy*, Supplementary Volume 1, Part 1. Reprinted in *Kant's Groundwork of the Metaphysics of Morals: Critical Essays*, ed. P. Guyer, 29-49. Lanham: Rowman & Littlefield, 1998.

Wood, A. E. 1999. *Kant's Ethical Thought*. Cambridge: Cambridge University Press.

ENDNOTES

¹ The ideas contained in this paper originated in a Spring 2000 graduate seminar on the *Groundwork* at Florida State University. I would like to thank all the seminar participants for their probing questions and useful contributions. This paper would not have reached its present form without the indispensable advice, criticisms, and constructive suggestions offered by David Brink, Barbara Herman, Pat Matthews, Al Mele, Dana Nelkin, Eric Watkins, and a referee for *Philosophy and Phenomenological Research*.

² References to Kant's works ("G" for *Groundwork of the Metaphysics of Morals*, and "CPR" for *Critique of Practical Reason*) are given parenthetically in the text, using the pagination of his *Gesammelte Schriften* (published by the *Preussische Akademie der Wissenschaften*, Berlin). Translations of relevant passages are quoted from Mary Gregor's editions of the *Groundwork* (Cambridge: Cambridge University Press, 1997) and the *Critique of Practical Reason* (Cambridge: Cambridge University Press, 1997).

³ Among those commentators who have contributed greatly to our understanding of the First Section, and who have done much to rectify misunderstandings on which much unwarranted hostility to Kant's project has been based, I count Allison 1990, Aune 1979, Herman 1993, Korsgaard 1989, Potter 1974, and Wood 1999. But none of these works, I claim, contains an accurate reconstruction of Kant's argument linking the good will with FUL in the *Groundwork*.

⁴ P1 should be distinguished from the following less restricted version thereof:

(P1*) An action has moral worth only if it is done from duty.

As a number of commentators have noticed, the difference between P1 and P1* is important. Kant does not accept that the actions of *any being* have moral worth only if they are done from duty. As Kant points out in the Second Section (at G 412 ff.), a perfectly good will, such as the will of God, is not bound by duty (because it automatically does what is good), and yet it is safe to presume that the actions of such a will can have moral worth. It is only in the case of humans (and, were there any, in similar creatures operating “under subjective limitations and hindrances”, i.e., sensuous desires or “inclinations”, that prevent them from *automatically* doing what is good) that actions must be done from duty in order to be morally good.

⁵ I am thinking primarily of Aune (1979), Korsgaard (1989, 1996b), and Allison (1991). My own view is that none of the ways of filling the gap proposed by these commentators succeeds. But a discussion of their proposals would take us beyond the scope of this paper, and is best left for another occasion.

⁶ Kant brings all three ideas (namely, absolute goodness, goodness without limitation, and unconditional goodness) together at G 437, in discussing “the idea of a will absolutely good without any limiting condition” (*die Idee eines ohne einschränkende Bedingung...schlechterdings guten Willens*).

⁷ This picture is at odds with the one presented by Aune (1979). Aune claims that Kant derives P1, not from (1) or even T1, but from P2! “The assumption on which [Kant] builds his argument,” Aune writes, “is that the moral value of an action depends entirely on the motive or ‘principle of the will’ from which it is done” (22). This assumption is P2. Taking this assumption to mean (tendentiously, in my view) that a human action has moral value if and only if its motive is “intrinsically admirable” (i.e., ensures that the action has moral value), Aune then claims that Kant argues that the only intrinsically admirable motive is the motive of duty. It follows directly that a human action has moral value if and only if it is done from (the motive of) duty, and this result is just a restatement of P1. But, as I’ve just argued, I do not believe that the text of the First Section supports Aune’s interpretation.

Wood (1999) correctly identifies (1) and T1 as two “principal claims that Kant seems to be making in the first three paragraphs of the First Section” (23), and identifies (one half of) P1 as Kant’s First Proposition. But Wood does not explain how these claims are supposed to be logically related. In particular, he mentions neither that T1 follows from (1) nor that P1 follows from (1) and (3).

⁸ Among those scholars who do not take Kant to exclude the possibility of overdetermination *ab initio* are Henson (1979) and Herman (1993). Jensen (1989, 161) goes further in arguing that Kant’s conception of indirect duty “implies that there will be occasions in which dutiful actions will be overdetermined by being done from the motive

of duty and from some other nonmoral motive or motives.” For criticism of Jensen’s argument, see Baron 1995, 170 ff.

⁹ Similar statements appear in various places in the *Critique of Practical Reason*. For example, Kant says that the moral law “excludes all inclinations from immediate influence on the will” and “an action that is objectively practical in accordance with this law, with the exclusion of every determining ground of inclination, is called *duty*” (CPR 80).

¹⁰ In saying that Kant excludes the possibility of overdetermination, I do not mean to suggest that Kant would have denied that an action could be done from the motive of duty while inclinations for (or against) the action are present. Nor do I mean to suggest that Kant would have denied that an action could be done from the motive of duty while inclinations are present that would have sufficed to produce the action in the absence of the motive of duty. What I *do* mean to suggest is that Kant would have denied the possibility of an action done *from* duty and inclination at the same time. For more detailed discussion of these matters, see Baron 1995, Chapter 5, and Herman 1993, Chapter 1. (I am indebted to a referee for helping me to clarify this issue.)

¹¹ Alternatively, Kant means by “maxim” what he later says “maxim” should be taken to mean, namely “subjective principle of volition” (in the footnote at G 400-401), so that the maxim of an action should be taken to be merely “the principle of volition in accordance with which the action is done” (G 399-400). But if so, then his statement of P2 at the

very beginning of (α) is incomplete. Rather than reading the way it does, the sentence stating P2 should read: “An action from duty has its moral worth not in the purpose to be attained by it but in the maxim in accordance with which it is decided upon *without regard for the ends that can be brought about by such an action.*”

¹² A proposition receives an asterisk when it plays a role in *some* argument of Kant’s but does not play a role in the version of *K* that I will construct on Kant’s behalf (see below).

¹³ According to Aune’s (1979, 23) reconstruction of Kant’s reasoning in (α) and (β), P2 “needs little discussion, because it is closely related to [P1]. If an action has moral value just when it is done for the sake of duty, its moral value must be dependent on the motive or, as Kant says, the ‘principle of the will’ from which it is done. But a principle of the will is simply a maxim in Kant’s technical sense of the term... Given this, Kant is entitled to say that the moral value of an action is owing to the maxim on which it is based rather than to its success in realizing some desired end or purpose.” As I hope the discussion in the text makes clear, Kant’s argument in (α) and (β) goes further than Aune recognizes. For Kant attempts to prove that the moral value of an action done from duty lies, not merely in a principle of the will, but in a principle of the will that is *a priori*.

¹⁴ Potter (1974) does not address this question.

¹⁵ It might be objected that Kant’s later acceptance of the claim that the rational will legislates *a priori* laws (at G 431) conflicts with premise (11) in my reconstruction of the argument for P3, namely with the thesis that *a priori* principles (including, presumably, *a*

priori laws) are not connected to the will as effects. This is an interesting concern, and worth pursuing. Here I can only gesture at an answer. As I see it, the relevant conflict is only apparent. For the idea of a will's legislating or giving a law is the idea of a will's being *the source of universal obligation*, which differs from the idea of a will's *bringing something into existence*. But it is only the latter idea that subsumes the idea of a law's being connected to a will as *effect*. (I am indebted to a referee for raising this issue.)

¹⁶ The phrase "universal law" is redundant, since Kant holds that universality is the essence of law. Thus, at G 408, he tells us that the moral law "is so extensive in its import that it must hold not only for human beings but for all *rational beings as such*".

¹⁷ It might be argued that the parallel between () and () is no more than superficial, since, as Kant himself claims, the aims of the respective sections in which they appear are so very different: the section in which () appears is devoted to discovering the supreme principle of morality from within "common rational moral cognition," while the section in which () appears contains an argument for the same principle within "philosophic moral cognition." It would seem to follow that if one is interested, as I am, in providing a proper reconstruction of the overall argument of the First Section, one should not help oneself to pieces of reasoning belonging to the Second, no matter how similar they might otherwise appear to be. My answer to this objection is that the different functions Kant takes his sections to be fulfilling do not render it illegitimate to appeal to an argument belonging to one section to bolster an argument belonging to another. For one thing, Kant does not say that his First Section relies *exclusively* on common rational moral

cognition in complete abstraction from philosophic moral cognition. After all, the First Section is entitled “Transition from common rational to philosophic moral cognition.” So we would expect to find some pure philosophical argumentation in the First Section. And, in fact, this is precisely what we do find at ()-(), passages in which many of the concepts invoked (such as formal vs. material principles, the determination of the will) are clearly part of philosophic, as opposed to mere common, moral cognition. So I find myself drawn to the view that both sections contain philosophical reasoning, parts of which may permissibly be used to shed light on other parts. (I am indebted to a referee for raising this issue.)

¹⁸ Kant says (at G 421) that “there is...only a single categorical imperative.” But he goes on to claim (also at G 421) that “all imperatives of duty can be derived from this single imperative as from their principle.” Since imperatives of duty are surely categorical (according to the definition of “categorical” at G 414), Kant must be supposing that there is more than one categorical imperative. (This is confirmed by his use of “categorical imperatives” (in the plural) at G 425.) This is all unnecessarily confusing. All that Kant means in saying that there is a single categorical imperative is that there is one categorical imperative from which all other categorical imperatives may be derived. Kant calls this imperative “*the* Categorical Imperative”. In using the definite article, he does not mean to suggest that categorical imperatives number no more than one.

¹⁹ To be precise, (22) follows from (21) and (OW) only on the assumption that (OW) is analytic. The reason is that it is only on the assumption of (OW)’s analyticity that we can

substitute “those maxims such that I (fictionally assuming that I know my will to be immediately effective) ought to will that every imperfectly rational being act on them” for “those maxims that every imperfectly rational being ought to act on” within the propositional attitude context generated by “says” in (21).

²⁰ To be precise, (23) follows from (22) and (OC) only on the assumption that (OC) is analytic. The reason for this is similar to the reason for requiring the analyticity of (OW) in the inference to (22) (see previous note).

²¹ For more extensive discussion of what both tests require, see Herman (1993, Chapters 6 and 7), Korsgaard (1985), O’Neill (1985), and Wood (1999, Chapter 3).

²² Whether K is sound is a question I leave aside for further discussion. I take it that the most controversial premises of K are these. First, there is (A1), the proposition that all actions done from inclination have no moral worth. For it might be thought, contrary to Kant’s own beliefs, that it is possible for actions to be motivationally overdetermined (i.e., done from inclination and from duty at the same time), in which case it would follow (contrary to A2) that (it is possible that) some actions done from duty have no moral worth. Moreover, not everyone agrees with Kant that the helping actions of a person are morally unworthy if performed from the motive of sympathy. Second, there is (14), the claim that only *laws* are connected with the will merely as ground and never as effect. For it might be thought that principles other than laws could, as Kant puts it, “outweigh [inclination] or at least exclude it altogether from calculations in making a

choice” (G 400-401). Finally, there is OW, the claim that the maxims that every imperfectly rational being ought to act on are maxims that I (fictionally assuming that I can make every imperfectly rational being act on them) ought to will that every imperfectly rational being act on. For, although it is difficult to see how OW might be false, Kant requires (see note 19) that OW be analytic, and it might be thought that OW is only contingently true, if true at all. Although these three assumptions are not nearly as obvious as Kant thought they were, I do not believe that they are easily dismissed. But I will not attempt to defend them here.