

How Are the Different Formulas of the Categorical Imperative Related?

IDO GEIGER

Ben-Gurion University of the Negev; Humboldt Universität zu Berlin

Email: geigeri@bgu.ac.il

Abstract

The article defends three claims regarding the relation between the different formulas of the categorical imperative. (1) On its prevailing reading, FUL gives different moral guidance than FH; left answered, this problem is an argument for adopting a competing perspective on FUL. (2) The prohibitions and commands of the formulas should be taken to be extensionally the same; but FKE adds a dimension missing from the others, gained by uniting their perspectives, namely, bringing the variety of moral laws into systematic unity. (3) The grammatically ambiguous phrase in *GMS*, 4: 436.9–10 claims that FA alone unites the other formulas in itself.

Keywords: formula of universal law, formula of humanity, formula of the kingdom of ends, equivalence of the formulas of the categorical imperative, system of moral laws, complete determination, love and respect, moral knowledge, moral guidance

The task of the *Groundwork of the Metaphysics of Morals*, Kant tells us in its Preface, is ‘nothing more than the search for and establishment of the *supreme principle of morality*’ (Kant 1996a: 47; *GMS*, 4: 392).¹ In the first section of the book, he indeed argues that the so-called formula of universal law (FUL) is this supreme principle (Kant 1996a: 57; *GMS*, 4: 402). It is presented again in the second section (Kant 1996a: 73; *GMS*, 4: 421). But now it is followed by two further formulations of the categorical imperative, the formula of humanity as an end in itself (FH) and the formula of autonomy (FA), as well as by variant formulations of FUL and FA, respectively, the formula of the law of nature (FLN) and the formula of the kingdom of ends (FKE). Kant then makes the following striking assertion:

The above three ways of representing the principle of morality [FLN, FH, FKE] are at bottom only so many formulae of the very

same law, and any one of them of itself unites the other two in it.
(Kant 1996a: 85; *GMS*, 4: 436)²

It is the purpose of this article to examine three of the important questions this assertion raises.

In the first two sections, I will examine the first part of the sentence, i.e. the claim that the three formulas are formulations of one and the same law. In section 1, I will ask whether the moral direction given by FUL and FH in particular is the same, as this claim seems clearly to suggest. Specifically, I will thus ask whether FUL and FH pass the same judgments on maxims (sections 1.1, 1.2) and whether they both command making the same ends our duties (section 1.3), or in other words whether the moral guidance they offer is extensionally the same. I will argue that, on its prevailing reading, FUL does not offer the same extensional moral guidance as FH and that attempts to respond to this difficulty are not fully successful. In light of Kant's assertion that the different formulas are formulations of the same law and the fact that one task of these formulations is to offer us moral guidance, this conclusion presents us with the following choice: either (1) find a systematic solution to the problem of different extensional guidance; or (2) reject the prevailing reading of FUL and adopt a different perspective, according to which the moral guidance of the formulas can indeed be extensionally the same (section 1.4).

In the second section I will focus on the moral guidance FA offers and specifically on Kant's claim that appraising action from the perspective of autonomy 'leads to a very fruitful concept dependent upon it, namely that of a *kingdom of ends*' (Kant 1996a: 83; *GMS*, 4: 433). I will claim that the laws FKE has in view should be taken to be extensionally the same as the laws enjoined by FUL and FH. But whereas FUL and FH address single agents and conceive of moral laws in the singular and determinately, autonomy and the kingdom of ends are regulative principles guiding a community of diverse moral lawgivers who each have as their legislative end the ideal of a comprehensive system of moral laws. The principal aim of this section is to explain what it means to be guided with diverse others by the regulative idea of a comprehensive system of moral laws.

I will then turn to the second part of the sentence and ask what precisely it claims and how consequently it should be translated. For as a matter of fact, the original German phrase 'deren die eine die anderen

zwei ... vereinigt' is ambiguous and might be read as making two logically very different claims: either (1) that *anyone* or *each one* of the formulas unites the other two in it; or (2) that *one* of the formulas unites the other two in it.³ So the question of how the sentence should be rendered in English (or other languages) is not merely one of translation and of little interest when reading Kant in the original.⁴ It is a matter of interpreting an important assertion. I will claim that the sense Kant seems to attribute to FA (the topic of section 2), as well as the manner in which it is explicitly introduced, strongly suggest that this formula has in view the very same laws enjoined by FUL and FH. But it also contains another very important dimension of moral direction missing from them, gained by uniting their two perspectives. There seems to be, however, no sense in which FUL and FH unite within them the other formulas, or clearly have in view a system of moral laws and a diversity of moral lawgivers as FA does. So I will conclude that what Kant is saying in the sentence under discussion is that FA is the one formula that unites the other two within itself (section 3).

1. Do FUL/FLN and FH Offer the Same Moral Guidance?

In the secondary literature, the sentence we are considering is very commonly referred to as the claim that the formulas of the categorical imperative are equivalent. Indeed, if the second part of the sentence claims that *each one* of the formulas unites the other two in it, then Kant does unequivocally claim that the formulas are equivalent in all respects. Each unites, has in it or contains the other two. Whatever one says or does the others too can say or do. It should be noted, however, that opting for this translation of the second part of the sentence might lead interpreters to read the first part of the sentence too as asserting the equivalence of the formulas. For the sake of clarity then it seems wise to consider the two parts of the sentence separately; I leave the questions of the sense and proper translation of the second part of the sentence for the last section of the article.

The first part of the sentence tells us that FUL, FH and FA are 'at bottom only so many formulae of the very same law' (Kant 1996a: 85; *GMS*, 4: 436). This part of the sentence is also sometimes taken to claim that the formulas are equivalent. Strictly speaking, however, it seems to claim something somewhat weaker. The three formulas of the categorical imperative are very closely related to each other, being different formulations of the self-same law. As the next paragraph makes clear, the claim is that the formulas present the categorical imperative from different perspectives: FUL presents the universal form of moral laws;

FH presents the matter or end of moral action, i.e. the capacities for rational and moral legislation that must never be violated but must rather be furthered; finally, FA has in view a ‘*complete determination of all maxims*’ (Kant 1996a: 86; *GMS*, 4: 436) (more on this later). Understanding how the three very different formulas are in fact formulations of the same law is not at all a trivial task. The fact that *Groundwork* I, which purports to begin with common rational cognition, discusses only FUL might suggest that it alone is present, in some form, to ordinary human understanding and that the derivation of the other formulas requires further premises available only from a metaphysical perspective. Kant himself is referring perhaps to this difficulty when he says that they are all formulations of the one moral law ‘at bottom’. If this is true, the formulas are not – at least not as they stand alone – equivalent in all respects. This suggestion is confirmed by the fact that in order to derive FH Kant introduces a new set of terms to explicate the workings of rational agency, specifically the notion of a will that determines itself to action in accordance with laws to attain ends it represents to itself. Indeed, Kant says explicitly that elucidating the connection between the universal law of morality and the will requires that we ‘step forth, namely into metaphysics’ (Kant 1996a: 77; *GMS*, 4: 426).⁵

1.1 O’Neill’s Interpretation of FUL and the Challenge of Counter-Examples⁶

This, however, is not a problem for our present concern. For the question I want to consider in this sub-section is whether the moral guidance given by FUL and FH is the same: do they prohibit acting on the same maxims and command making the same ends our own? In other words, are their directions extensionally the same? That they should be does seem to follow from the fact that they are formulated as action-guiding principles that respectively have in view the form and matter of moral laws and the claim that the formulas are all formulas of the one moral law. As different formulations of one and the same law, FUL and FH should give us the same moral guidance extensionally. But this is just the sense usually given to the equivalence claim.⁷

Now according to FUL, maxims that cannot be universalized without contradiction in conception or in the will are morally impermissible. FH too contains a prohibition. It is impermissible to make the humanity of a person a mere means to our ends. But we cannot just assume that the two formulas prohibit acting on the same maxims, though Kant does hold

that both prohibit the maxims in the four examples he gives.⁸ Rather, this is something that has to be shown.

On its most prevalent interpretation, developed and defended most fully by O'Neill, FUL presupposes no moral knowledge. More specifically, the terms we use to formulate our maxims (to take one's own life, to borrow money, etc.) are taken to be morally neutral. Testing whether a maxim can be universalized without contradiction in conception or in the will decides its moral status.⁹

Notoriously, however, readers have come up with many counter-intuitive false positive and false negative results of the universalization test. I will not attempt to give a complete list of references or a long list of reasonably convincing examples.¹⁰ For our present purpose, it would be useful rather to circumscribe three kinds of examples that do yield some troubling counter-intuitive results. First, very many of the false negatives are so-called maxims of coordination, i.e. maxims that comprise ends that seem to express no immoral intent, but nevertheless lead to contradictions in conception when universalized, for example: (1) queue up early and be the first in and out of the bank, in order to get to work on time; (2) hold the door open for others, in order to express consideration for them. Second, many false positives are maxims that do seem to express an immoral intent to transgress a strict duty. They thus should lead to a contradiction in conception; but they do not, because the maxims include an excess of detail, for example: (1) in order to pay a debt, make a false promise to someone who is irremediably gullible, or (2) to someone who is painfully shy and cannot bear to express any suspicion. The third group of examples consists of so-called maxims of violence. These express the inner and clearly immoral modes of thought which guide those brazenly intent on harm or violence and seem clearly to be transgressions against strict duties, for example: (1) manipulate, (2) deceive, (3) threaten or (4) physically coerce the weakest and most helpless, in order safely to gain advantage.

How does FH fare in relation to these problems? In the first case the maxims seem not to make the humanity of a person a mere means to the agent's ends and in the latter two cases they clearly do. FH does not seem to be vulnerable to these counter-examples. The prohibitions of the two formulas are not then the same.

Clearly, the way to avoid the conclusion that FUL and FH do not prohibit the same maxims is to find a systematic solution to the false positive and false negative results of the universalization test. There are two general

strategies for doing this: (1) remain committed to the assumption that employing the test presupposes no moral knowledge and find plausible formal strictures on what maxims are tested; these strictures should systematically exclude the problematic counter-examples; (2) accept that agents employing the test possess considerable moral knowledge.

But this latter line of response, to which we will return below (1.4), is blocked to the prevailing reading. For it, the assumption that FUL presupposes no moral knowledge is central. The thought here is that the test must presuppose no moral knowledge if it is to yield objective results. O'Neill makes this point very clearly: 'if practical reasoning is to show why we should adopt some rather than other principles as maxims, setting prior limits on which principles are to be adopted as maxims would beg questions'.¹¹

O'Neill offers a partial answer to the problem under consideration by setting formal limits on what counts as a maxim, specifically by characterizing maxims as '*underlying principles or intentions by which we guide and control our more specific intentions*'.¹² Much of the detail of our intended course of action is not, on this view, part of the maxim; and the same holds true of very general descriptions of it. So some very general and specific descriptions of our intentions would be ruled out by this very plausible stricture. But detailed maxims leading to false positives would not be ruled out when the problematic details are part of the guiding principle.¹³ For then, on O'Neill's view, they are to be tested: 'an agent's maxim in a given act must incorporate just those descriptions of the agent, the act and the situation upon which the doing of the act depends'.¹⁴ My making a false promise might really depend upon finding a gullible enough person to deceive. Nor does this stricture rule out the maxims of coordination or violence that do describe the underlying principles guiding our action. More generally, it is hard to see how any formal stricture on what counts as a maxim is going to rule out familiar, action-guiding intentions of the sort given above.

1.2 O'Neill's Equivalence Argument

As I said in section 1.1, O'Neill is the most thorough and influential proponent of the canonical interpretation of FUL. But she is also the author of a sustained argument for the equivalence of FUL and FH. It is particularly instructive to examine this argument.

I emphasized above O'Neill's claim that tests of universalizability, if they are to be objective, must take neither the formulation of maxims nor the

choice of maxims to be tested to depend upon prior moral knowledge. This claim is not made explicitly in O'Neill's paper on FUL, upon which the equivalence argument most directly draws. It is, however, implicit there in her contrast between Kant's emphasis on autonomous rational agency and heteronomous tests of universalizability, which 'construe moral acceptability as contingent upon the natural phenomena of desire and inclination, rather than upon any intrinsic or formal features of agents or their intentions'.¹⁵

In the same paper, O'Neill speaks of maxims of deception and coercion as only one prominent example of maxims that allegedly fail Kant's contradiction in conception test.¹⁶ But in her equivalence paper, when she attempts to contend with purported counter-examples to FUL, she deals solely with the case of brute or coercive violence.¹⁷ There are two significant problems with this. First and obviously, this is not a complete answer to the problem of counter-examples to FUL that do not raise a problem when examined under FH. It is understandable that O'Neill does not take on the problem of detailed maxims, for her notion of underlying principles is supposed to deal with it successfully. (I claimed in section 1.1 that it is not a fully adequate response.) But she does not deal with the problem of maxims of coordination.

Second and more interestingly, when contending with the problem of violence, she is apparently thinking of agents who conceive of their actions as employing the destruction or constraint of agency as a means to their ends.

There is a palpable contradiction in the thinking of an agent who adopts a maxim of assault or murder, or of duress and intimidation, which aims to destroy or undercut at least some other's agency, yet (tries to) will the same maxim as a universal law. Agents cannot coherently (nor honestly) assume that the agency of those whom they *plan* to destroy or damage can *already* be discounted! It is only after a killing that its victims are no longer agents; before the killing they are agents and must fall under the scope of FUL; victims even of minor coercive violence are evidently agents before and after the violence, which cannot be willed as a universal law because it aims to undercut agency, at least for some time.¹⁸

There is indeed a contradiction between willing the destruction or restriction of agency as a universal law and at the same time willing your

own unrestricted agency. The problem is that the contradiction depends upon conceiving of the action as a coercive act of violence and an assault on humanity. This seems clearly to attribute moral knowledge to agents. The contradiction might very well not arise otherwise. The maxim of the confidence-man might plausibly be to employ his unusual ability to inspire trust in others in order to deceive them. His maxim would not fail the universalization test.

The problem is also visible in the attempt to show that what is ruled out by FH as a maxim that treats another merely as means cannot be universalized without contradiction in conception. O'Neill interprets FH and the notion of respect for persons as concerned paradigmatically with deceit and coercion and thus with consent and dissent.¹⁹ The heart of this part of her equivalence argument is the claim that treating others as mere means is acting without allowing them to consent to your action and possibly participate in it. They cannot be willed as universal laws without contradiction in conception, precisely because others are not given the opportunity to consent.²⁰

The problem, of course, is that so conceiving of the universalization tests is attributing to agents considerable moral knowledge. Maxims of coordination, for example, are precisely cases in which the consent of others seems to be irrelevant. They typically make use of known patterns of behaviour or reasoning of other agents. But acting on this information without obtaining the consent of those affected by our action does not count as either deceit or coercion. I rush to the bank without securing anyone's consent in order to be the first in line (and knowing there are probably others with the same aim). But I am not deceiving anyone. In this case, we would like to say, no consent is needed. Indeed, others have no moral ground for dissent. This, presumably, is why acting in this way is not forbidden by FH; and so conceived, these maxims would not, on O'Neill's view, be ruled out by FUL either. But to conceive of the action in this way is to attribute considerable moral knowledge to agents.

In the examples of morally irrelevant detail, I am using another – indeed, quite shamefully – as a mere means to my ends. But has the mortally shy person not given her consent to lend me the money, even though she must know I will not pay her back? Does the irremediably gullible person, to whom I and others have lied many times before, not give his consent? We would like to say that, although they have given me the money, in an important sense they have not expressed their consent, that I am clearly using them as means to my ends, indeed precisely because they cannot as

a matter of fact actually express dissent. But to say this means that we know that there is a sense in which they have not freely consented to lend me the money, though they have given it to me and do or should know full well the outcome. Again, we would have to conceive of the deed in just this way to be ruled out by FH and by FUL as well. The problem, in sum, is that to determine what precisely constitutes deceit or coercion and when precisely consent is needed and has been given or dissent expressed seems clearly to demand considerable moral knowledge.

O'Neill claims that tests of universalizability must not assume prior moral knowledge. But in arguing for the equivalence of FUL and FH she apparently assumes that agents do possess such knowledge. It is interesting to see that her reading of FH strives to strike a middle course between a purely formal approach to the second formula and an empirically concrete approach. Specifically, we saw that O'Neill interprets the notion of respect as concerned paradigmatically with deceit and coercion and thus with consent and dissent. The central task of her FH paper then is to interpret the notion of consent. She argues that the actual (empirical) consent of agents is often not 'genuine, morally significant consent'.²¹ But a purely formal approach to consent, the consent of hypothetical perfectly rational agents, risks the clearly unwelcome conclusion that 'treating others as persons should even sometimes be a matter of overriding what others as we know them actually choose'.²²

O'Neill interprets the moral notion of consent as having in view the very possibility of consent. In other words, the question is whether we have made it possible for actual others to consent or dissent to the actions that affect them and that we propose to carry out. O'Neill states very clearly that an 'account of genuine consent must then show how the morally significant aspects of plans, proposals and intentions are picked out as candidates for consent'.²³ Indeed, O'Neill's stated aim is to give an account that views persons 'as particular men and women with limited and determinate capacities to understand and to consent to proposals for action'.²⁴ This seems clearly to demand attributing considerable moral knowledge to agents.

In conclusion then, I would argue that O'Neill's reading of FH as well as her equivalence argument presuppose agents who have acquired considerable moral knowledge. But her stated position appears to be that universalization tests must not assume moral knowledge. This is the source of tension between her readings of the first two formulas and explains why FUL (on her reading) is and FH is not vulnerable to counter-examples.

1.3 *Maxim Permissibility and the Comprehensive Derivation of Duties*

The second argument showing that on the prevailing reading FUL and FH do not offer the same moral directives is simpler, yet no less significant. According to the prevailing reading, FUL is taken to be a procedure or test that determines whether acting on a maxim is morally permissible or impermissible. If a maxim cannot be universalized without contradiction, it is impermissible; if it can – it is permissible. This is not insignificant. Learning that making false promises to escape financial straits or committing suicide when sunk in deep depression are morally impermissible teaches us that we ought to keep our promise and preserve our lives in these circumstances.²⁵ But even if the test were infallible in its pronouncements, it would fall far short of the suggestion that ‘all imperatives of duty can be derived from this single imperative as from their principle’ (Kant 1996a: 73; *GMS*, 4: 421; cf. Kant 1996a: 79–80, 86; *GMS*, 4: 429, 436).²⁶ Again, this is so not because the test is wholly uninformative about our moral duties, but because it is less than fully informative. We cannot derive all duties from the test, first, because finding out that universalizing a maxim does not lead to contradiction might tell us that acting on it is permissible. (I owe to Allen Wood the point that Kant never uses the formula to show the permissibility of a maxim.) But it does not tell us whether it is in fact morally obligatory to do so; for example, the test might pronounce that I am permitted to expend time on developing my natural talents, but that does not tell me whether it is my duty to do so – and Kant thinks it is a wide duty. Second, it relies on cases being brought before its tribunal as they come up, so to speak. It does not offer us a complete inventory of maxims we morally ought to act upon and ends we ought to make our own, as Kant seems at least to suggest the first two formulas do.

These points are very often noted by readers and a principal reason for turning specifically to FH to make up for the shortcomings of FUL. First, the prohibition on making the humanity of a person a mere means to our ends does seem to rule out certain types of action such as manipulation, deceit and coercion; this is a more general prohibition than proscribing acting on a specific maxim that is actually being considered by an agent in a given situation. In this respect, the scope of moral insight FH offers is wider. Second, FH has a positive side telling us it is our duty to make the humanity of persons our end. In the *Doctrine of Virtue*, Kant claims that there are two very general ends that are duties, ‘*one’s own perfection and the happiness of others*’ (Kant 1996a: 517; *MS*, 6: 385). My own moral perfection and the happiness of others are precisely what I make into ends when I treat humanity in my own person and in other persons

as an end. Both general ends entail positive duties. Granted, FH offers us only very general directives: we must never limit and indeed must further the ability of others to choose their ends and act to fulfil them; and we must cultivate our own agency and specifically strive to fulfil our own moral vocation. But on the prevailing reading, FUL does not offer any corresponding broad direction.

O'Neill has proposed to respond to this problem by claiming that where 'a maxim is universalizable but the maxim of rejecting it is not, the first maxim is one of duty and to act out of it would not merely not be morally unworthy but morally worthy'.²⁷ The problem with this solution is that it does not seem to be true to the manner of thought of ordinary agents, which Kant purports to describe. It is not at all obvious that a person who very much wants to act on a certain maxim and finds out that it is permissible to do so would stop to examine whether it is impermissible not to act as he desires. But Kant says of the categorical imperative under its first formulation that 'it would be easy to show how common human reason, with this compass in hand, knows very well how to distinguish in every case that comes up what is good and what is evil, what is in conformity with duty or contrary to duty ... without in the least teaching it anything new' (Kant 1996a: 58; *GMS*, 4: 404). Indeed, it seems that in order to discover we are bound by a wide duty, we would have to test the negation of the comprehensive disjunction of ways in which the duty can be fulfilled and find out that this is impermissible.

Even if we ignore this problem, knowing that we ought to act on certain maxims is dependent on the maxims actually occurring to agents, who then test them and their negation. So, for example, people who never ask themselves whether it is morally permissible to develop their talents (or forbidden to neglect them entirely) would not know it is not only permissible but indeed their duty. Moreover, this answer does not tell us what action types are generally prohibited and what ends we ought to adopt as our own. This indeed is a point O'Neill acknowledges:

The Categorical Imperative provides a way of testing the moral acceptability of what we propose to do. It does not aim to generate plans of action for those who have none.²⁸

In conclusion, the fact that, on the prevailing reading, FUL as a test for the permissibility of maxims that an agent is considering does not offer the same guidance as FH constitutes a second, very significant challenge to this reading.

1.4 *FUL, the Presupposition of Moral Knowledge and Equivalence*

Is there a different strategy of reading *FUL*, one that might not be vulnerable to the problems of false negative and false positive results as well as to the problem that *FUL* is a test of permissibility only and so to the problem that the moral guidance given by *FUL* and *FH* is different? Though I think still a minority (at least in the English-speaking world), there are noteworthy exceptions to the standard reading, who do take agents who employ the universalization test to rely upon prior moral knowledge. Best known is Herman's claim that agents have a grasp of 'rules of moral salience':

Acquired as elements in a moral education, they structure an agent's perception of his situation so that what he perceives is a world with moral features. They enable him to pick out those elements of his circumstances or his proposed actions that require moral attention.²⁹

Others have followed Herman.³⁰ It should be clear that for such readings the challenges discussed above need not arise. They would not arise if the notion of rules of moral salience or the presupposition of moral knowledge more generally can be shown to rule out the sort of problematic maxims that plague the prevalent reading of *FUL*. They might also ensure that the right maxims are tested at the right time, and so would not face the challenge that relevant moral questions will not be raised by agents. Such readings can assume that agents generally know what is morally required of them.

One such reading is Wood's, who advocates viewing the different formulations of the categorical imperative as a system and specifically viewing the first formal formulations (*FUL/FLN*) as underwritten by the specification of a substantive conception of moral value through *FH*.³¹ And focusing not on the content but on the form of practical knowledge, Engstrom, who argues in detail for the equivalence of all three formulas, holds that thinking in accord with the formal demands of practical knowledge presupposes the sort of moral knowledge ruled out by the prevalent reading.³² He claims, for example, that the very application of *FUL* 'involves the requirement that others' happiness be deemed *equal* to one's own in basic goodness'.³³ It further presupposes knowing how to distinguish 'between elements of happiness that are necessary to it ... and elements that are by comparison arbitrary and contingent ... other objects of pursuit, such as those figuring in innocent forms of competition, are not only contingent but not even objects of practical knowledge'.³⁴

Moral agency presupposes ‘a body of practical knowledge that is universally self-consistent and always implicitly present’.³⁵ The distinction between innocent and culpable competition solves the problem of maxims of coordination, because they are typically at bottom fair competitions for limited means. Moreover, the problem of overly specific maxims yielding different results than their more general counterparts would not arise in a ‘body of knowledge that is universally self-consistent’. The consistency requirement also rules out maxims of violence. Finally, the body of practical knowledge clearly includes knowledge of general moral prohibitions and mandatory ends.

That presupposing moral knowledge can answer the challenges is also a point I have made explicitly in raising the following possibility:

Kant’s agents have informal knowledge of what moral laws bind them and what actions these laws command or forbid in almost all everyday situations. They possess this knowledge *without employing any formal universalization test*.³⁶

For such a reading, whatever its other faults, the problems of false negatives and false positives do not arise.³⁷ Furthermore, possession of such knowledge makes FUL more than a test of permissibility. For on this reading, FUL is not a procedure for the objective derivation of moral laws.

In conclusion, as the last examples show quite clearly, for readings that attribute considerable moral knowledge to agents, the challenges facing O’Neill’s reading of FUL do not arise. It should be noted, of course, that on these readings FUL is not taken as a test that establishes the objective validity of substantive moral laws. This result might be held by many readers to be philosophically too dear, for this is precisely what many hope FUL provides. As I suggested above, plausible formal strictures on what maxims are put to the universalization test might be found by defenders of the prevalent approach. I hope it is in any case not contentious to conclude this section by saying that showing that FUL and FH can offer the same moral guidance should be recognized as a constraint on any interpretation of these formulas.³⁸ This though means that it is not legitimate to solve problem cases posed for FUL by appealing to FH and Kant’s claim that the different formulas are formulas of one law.

2. Do FA/FKE Offer the Same Moral Guidance as FUL/FLN and FH?

FA and the notion of autonomy generally are of the greatest importance for Kant’s theory of morality. Indeed, he says explicitly that the ‘principle

of autonomy is the sole principle of morals' and that a categorical imperative 'commands neither more nor less than just this autonomy' (Kant 1996a: 89; *GMS*, 4: 440).

But what precisely does Kant mean by 'the idea of the will of every rational being as a will giving universal law'? First, though Kant is clearly offering formulations of the moral law that categorically binds all rational agents, *FUL* and *FH* are phrased in the second person singular. That we are talking about the way all rational agents are to legislate – and indeed, with a view to the diversity of others – is made explicit in *FA*. *FA* emphasizes that moral legislation requires that we all think of the ends 'each may set himself' (Kant 1996a: 83; *GMS*, 4: 433). Second, the first formulations are immediately applied to specific situations. In contrast, *FA* makes explicit the fact that we are talking about a general principle of legislation for a system of laws. Third and very importantly, *FA* clarifies that the moral law is an idea of reason. In sum, the notion of autonomous legislation is the regulative idea guiding the ongoing activity of a moral community legislating with a view to human diversity a comprehensive system of moral laws.

Focusing on the question of the moral guidance of *FA*, I claimed above that the first part of the sentence we are considering strongly suggests that the different formulas of the categorical imperative give us the same moral guidance. We should read the claim, I said, as asserting (amongst other things) that the different formulas all prohibit acting on the same maxims and command making the same ends our own. In this section, I want to make this claim more precise. I will suggest that *FA* and, more emphatically, its variant *FKE* should be read as having in view the very same maxims and ends enjoined by *FUL* and *FH*. Extensionally, the directives of the three formulas should be taken to be the same. So when Kant claims that the difference between the formulas is subjectively rather than objectively practical (Kant 1996a: 85; *GMS*, 4: 436), we should take him to mean that there is no difference in what the formulas command but in how they present our moral duties.³⁹ But *FA* directs us to acquire a deeper and more comprehensive insight into the moral laws that bind us. This implies the point to which we will turn in the next section: the different formulas of the categorical imperative offer the same moral direction extensionally, but they are not in all respects equivalent.

The first, very important thing to note when focusing on the question of moral guidance is that Kant does not employ *FA* in deriving or appraising particular duties, neither in the *Groundwork* nor elsewhere in his

published writings.⁴⁰ Nevertheless, he does say that appraising action from the perspective of autonomy ‘leads to a very fruitful concept dependent upon it namely that of a *kingdom of ends*’ (Kant 1996a: 83; *GMS*, 4: 433). For our purposes then it is the variant FKE that is of particular importance.

Now a kingdom, Kant explains, is ‘a systematic union of various rational beings through common laws’ (Kant 1996a: 83; *GMS*, 4: 433) and a system is an ordered and comprehensive whole of laws organized under an *a priori* idea (Kant 1997: 691; *CPR*, A832–3/B860–1). Thus to appraise our actions from the perspective of autonomy is to be guided by a comprehensive conception of our various moral duties. It is to act not only from respect for a determinate moral law or value but also at the same time to seek a broader and more detailed vision of ideal moral community and a deeper understanding of how our many and various duties and values are related to each other. FUL and FH conceive of moral laws in the singular and determinately; and they address single agents. In contrast, autonomy and the kingdom of ends are the regulative ideals guiding a diverse community of moral lawgivers. To be guided by FKE then is to be guided with others by the regulative idea of the systematic unity of moral laws. In this precise sense, and again anticipating the claim of the next section, FA is equivalent neither to FUL nor to FH, although the three formulas should all be taken to command acting on the very same moral laws.

How, more concretely, might the guidance of FKE differ from the determinate directions of FUL and FH? Kant does not tell us. In the *Groundwork*, he is apparently concerned more with the metaphysical content of the idea and goes on to develop the notions of being a member in a kingdom of ends and of the dignity of such members. But here are some thoughts about the unique perspective FKE might offer.

2.1 Broad-Minded Legislation

Kant does not develop in the *Groundwork* the thought that a broad view of moral legislation must take into account the differences between persons. In a somewhat confusing sentence, he tells us first to ‘abstract from the personal differences of rational beings as well as from all content of their private ends’ (Kant 1996a: 83; *GMS*, 4: 433), presumably setting aside only what is personal or private, i.e. those ends which others cannot in principle further. He goes on to require that we take into account ‘the ends of his own that each may set himself’ (*ibid.*).⁴¹ What might he mean?

We can perhaps approach this question by recalling that one problem for the prevalent reading of FUL is that nothing guarantees that the morally pertinent questions arise for an agent. The alternative approach can deal with this problem by attributing to agents knowledge of what calls for moral consideration in any situation. This implies that we are always to think in what Kant calls a broad-minded way of what morality requires: ‘To think in the position of everyone else’ (Kant 2000: 174; *CPJ*, 5: 294). Most fundamentally, to think of others as ends in themselves is to make their happiness our end. More specifically, we are to think of the great variety of ways others might seek happiness or find themselves in need of help. We might take Kant to be directing us to think broadly and concretely about the injunction to seek a ‘positive agreement with *humanity as an end in itself*’ (Kant 1996a: 81; *GMS*, 4: 430). He would be then making explicit something merely implied by FH. As we shall see in the next section, there is good reason to take Kant to be making explicit in FA what remains implicit in FUL and FH.

In a short discussion in the *Metaphysics of Morals* Kant explains why the question of the detailed diversity of moral obligations to others is not part of pure ethics, but adds that they would belong to a complete presentation of the system of ethics. Here Kant asks what duties human beings have towards one another ‘with regard to their *condition*’ (Kant 1996a: 584; *MS*, 6: 468).

These (duties of virtue) do not really call for a special chapter in the system of pure ethics; since they do not involve principles of obligation for human beings as such toward one another, they cannot properly constitute a *part* of the *metaphysical* first principles of a doctrine of virtue. They are only rules modified in accordance with differences of the *subjects* to whom the principle of virtue (in terms of what is formal) is *applied* in cases that come up in experience (the material) ... Nevertheless, just as a passage from the metaphysics of nature to physics is needed – a transition having its own special rules – something similar is rightly required from the metaphysics of morals: a transition which, by applying the pure principles of duty to cases of experience, would *schematize* these principles, as it were, and present them as ready for morally practical use. How should one behave, for example, toward human beings who are in a state of moral purity or depravity? toward the cultivated or the crude? toward men of learning or the ignorant ... How should people be treated in accordance with their differences in rank, age, sex, health, prosperity or poverty and so forth? (Kant 1996a: 584; *MS*, 6: 468–9)

A detailed system of moral obligations would have to include these and other differences between human beings.

2.2 *On the Relations between Different Duties: Love and Respect for Others*

How else might the guidance of FKE be distinct? Consider the following passage from the *Metaphysics of Morals*.⁴²

Love and respect ... can be considered separately (each by itself) and can also exist separately ... But they are basically always united by the law into one duty, only in such a way that now one duty and now the other is the subject's principle, with the other joined to it as accessory. (Kant 1996a: 517; MS, 6: 448)

Kant, in this passage, is speaking about the wide duty of love or active benevolence owed to others and the strict duty of respect for others. They are, Kant *tells* us, distinct duties. They are not analytically related and we can consider them separately. Indeed, in the *Groundwork*, the fourth examples following the introductions of FUL and FH deal with the duty of benevolence – and they do so without drawing our attention to the close connection between benevolence and respect. However, in the passage we are now considering, Kant tells us that the duties are always connected. They are connected by virtue of the fact that, as Kant colourfully puts it, the ‘principle of *mutual love* admonishes them constantly to *come closer* to one another; that of the *respect* they owe one another, to keep themselves apart’ (Kant 1996a: 568–9; MS, 6: 449; see also 1996a: 585; MS, 6: 470). Indeed, we learn later that the unattainable ideal of a perfect union of love and respect in friendship is the perfect moral relation to another. And the ideal of becoming a ‘*friend of human beings*’ is very well described as creating the kingdom of ends. In such an ideal universal friendship love and respect fuse into one, for it adds to love of human beings ‘thought and consideration for the *equality* among them, and hence the idea that in putting others under obligation by his beneficence he is himself under obligation, as if all were brothers under one *father* who wills the happiness of all’ (Kant 1996a: 587; MS, 6: 472–3).⁴³

The example Kant then gives is very illuminating.

So we shall acknowledge that we are under obligation to help someone poor; but since the favor we do implies that his well-being depends on our generosity, and this humbles him, it is our duty to behave as if our help is either merely what is due him or

but a slight service of love, and to spare him humiliation and maintain his respect for himself. (Kant 1996a: 568; *MS*, 6: 448–9)

To see ourselves as bound by the duty of benevolence is to acknowledge, amongst other things, that we are under obligation to help the poor. (Recall that Kant specifically mentions the human condition of poverty in the passage we considered in the previous section.) But if in helping a poor man we present ourselves as his superiors, who have generously chosen to be his benefactors (the duty, recall, is a wide one), we are likely to humiliate him, wounding his self-esteem. The duty of respect for others, Kant tells us in the next section, is ‘strictly speaking, only a negative one (of not exalting oneself above others)’ (Kant 1996a: 568; *MS*, 6: 449–50). So in helping a poor man we must take great care not to present ourselves as in any way superior to him. Kant is suggesting then that in order to succeed in fulfilling our wide duty of benevolence to others we must take great care not to fail to obey our narrow duty of respect for them. Specifically, Kant seems to be suggesting that one way to do this would be to make the other feel that our help is no great matter at all (‘a slight service of love’) rather than a grand act of generosity. Another way to do so might be to act as though we are simply obeying a strict duty, doing what anyone ought to do for another (‘our help is merely what is due him’). In other words, Kant is suggesting we might behave as though our wide duty of love were in fact a strict duty ‘the observance of which does not result in obligation on the part of others’ (Kant 1996a: 568; *MS*, 6: 448; cf. Kant 1996a: 576; *GMS*, 4: 458). The more personal relation definitive of wide duties is presented as impersonal. The important general point is that we would fail to obey the wide duty of benevolence, which we acknowledge and indeed actively adopt as our end, if we do not take care to obey *at the same time* our strict duty of respect for others. Respect is here a success condition of benevolence and in this way serves as an accessory to it.

Indeed, this passage might lead us to think that the moral failing of the man described in the fourth example of FUL in the *Groundwork* might just be his blindness to the close connection of respect for the rights of others, the moral duty of respect and the duty of benevolence. Kant has him say: ‘I shall take nothing from him nor envy him; only I do not care to contribute anything to his welfare or to his assistance in need!’ (Kant 1996a: 75; *GMS*, 4: 423). We might read the first part as referring to respect for the rights of others.⁴⁴ Kant says explicitly that the ethical duty not to exalt yourself over others is ‘analogous to the duty of right not to encroach upon what belongs to anyone’ (Kant 1996a: 568; *MS*, 6: 449–50). The man

recognizes his juridical duty but perhaps not the analogous ethical duty, commanding us to check our self-esteem by recognizing the equal dignity of less fortunate human beings. Furthermore, Kant's mention of envy does not merely add colour to his example. It has a systematic conceptual connection to benevolence. Indeed, as we learn in the *Metaphysics of Morals*, envy is one of the vices directly opposed to the duty of benevolence (Kant 1996a: 576; *MS*, 6: 458–9). It is not enough, however, not to envy others, for envy is a vice that has duty opposed to it as a contrary. In terms of FH, this is just the difference between a merely 'negative and not a positive agreement with *humanity as an end in itself* (Kant 1996a: 81; *GMS*, 4: 430).

Finally, envy is but one example of the category of vices opposed or contrary to wide duties, which Kant develops in the Doctrine of Virtue (Kant 1996a: 576–9; *MS*, 6: 458–61) with the categories of vices which violate strict duties to oneself (Kant 1996a: 546–59; *MS*, 6: 421–37) and to others (Kant 1996a: 581–3; *MS*, 6: 465–8). These categories contain further examples of the sort of insight we can gain from considering moral duties and transgressions from a systematic perspective, though here we are no longer talking about the relations between different duties. I emphasize finally that all the insights gained by thinking of our various moral duties systematically have both a theoretical and a practical aspect. We gain a deeper insight into the moral laws that bind us and this insight can have a decisive influence on the way we act.

One last thought: moral legislation is always a matter of individual choice and action. This is the very idea of being autonomous, of legislating for oneself. But some moral ends cannot be attained unless we act coordinately with others. Indeed, the greatest moral goods (and evils) of this world demand joint action. Autonomy does not contradict acting in community with others. Such an ideal moral community would be worthy in the fullest sense of being called a kingdom of ends.

3. Does Each of the Three Formulas Unite the Other Two within Itself?

I claimed in the previous section that we should take the moral laws that FUL, FH and FA command to obey to be extensionally the same. But I also claimed that the latter adds to the former the regulative demand to think about the ideal of the systematic and comprehensive connections between all these duties. This means that the different formulas are not intensionally equivalent in the moral guidance they offer. Put in Kant's terms, it means that it is not the case that each one of the formulas unites

the other two in it. If this were the case, the guidance of the three formulas would not only be extensionally the same, but identical in all respects, for then both FUL and FH would unite within them and so contain FA (as well as each other).

The claim that FA unites within it FUL and FH finds strong support in the way it is introduced. For Kant says explicitly that FA follows from FUL and FH jointly. This is the way the first of the several formulations of FA is presented:

[T]he ground of all practical lawgiving lies (in accordance with the first principle) *objectively in the rule* and the form of universality which makes it fit to be a law (possibly a law of nature); *subjectively*, however, it lies in the *end*; but the subject of all ends is every rational being as an end in itself (in accordance with the second principle); from this there follows now the third practical principle of the will, as supreme condition of its harmony with universal practical reason, the idea of *the will of every rational being as a will giving universal law*. (Kant 1996a: 81; *GMS*, 4: 431)

We might say, in other words, that FA is obtained by uniting FUL and FH. This would explain why Kant sees no need to test the third formula on his four examples and says that the previous tests ‘can all serve for the same end here’ (Kant 1996a: 82; *GMS*, 4: 432n.).⁴⁵ Note too that Kant goes on to say that FLN can be thought under the category of unity; FH under plurality; and FKE under allness or totality (Kant 1996a: 86; *GMS*, 4: 436). Whatever the precise sense and significance of this claim, we know that Kant holds that the third category ‘always arises from the combination of the first two in its class’ (Kant 1997: 215; *CPR*, B110).⁴⁶ A little later Kant suggests again that the third formula follows from the first two (Kant 1996a: 87; *GMS*, 4: 438).⁴⁷ Finally, Kant’s claim that the third formula unites the other two of itself (*von selbst*) might just mean that it is nothing but the product of this unification – this is just what it is or has in it.⁴⁸

Explaining in what way precisely FA unites the other two formulas in it is not an easy task. In the one long sentence first introducing it we have nothing like a formal argument and this makes it difficult to reconstruct Kant’s line of thought. But here is what I think Kant might have in mind. The objective ground of moral lawgiving, he says, is the universality of a law. The subjective ground of moral lawgiving is the end of moral action,

i.e. humanity in our own person and in the person of others.⁴⁹ In other words, moral action is directed at rational beings and specifically at their capacity for moral legislation.

Implicit in the notion of universal moral laws as the objective ground of morality is the notion of rational legislation. The focus of the first formula is the universal form of the law – the universal form of the law is evidence of its objective authority. But the law, of course, is the product of rational legislation. It is reason alone that gives universal laws. But to turn our attention to reason as the capacity for rational legislation reveals that we are not talking about single distinct laws but about an ongoing activity governed in all its decrees by one principle. It is, furthermore, the characteristic demand of reason to seek comprehensive unity in knowledge and practical legislation. Thus the idea of a comprehensive system of laws, merely implicit in FUL, is made explicit in FA.

The subjective ground of action is the perspective of rational agency; and rational agents, Kant stresses, direct their action at ends they represent to themselves. FH teaches us that moral action is directed towards rational nature or rational beings. As we saw in the previous section, the great variety of ways in which others might seek their happiness or meet with difficulty remains merely implicit in the command to seek a ‘positive agreement with *humanity as an end in itself*’ (Kant 1996a: 81; *GMS*, 4: 430). This diversity, I claimed above, is made explicit in FA, most clearly in the demand (following the introduction of FKE) to take heed of ‘the ends of his own that each may set himself’ (Kant 1996a: 83; *GMS*, 4: 433).

So joining the two formulas gives us a diversity of rational beings, each with his or her own conception of happiness, who are to obey morality in all their legislation. This is the idea of the legislation of a system of moral laws for a variety of human beings.

In conclusion then the way in which Kant introduces FA and the sense attributed to it strongly suggest that he is saying that FUL and FH are united within FA. Indeed, he is perhaps further suggesting that only by uniting the former formulations can access be provided to the new dimension the latter adds. But I see no sense in which FUL and FH unite the other two formulas within themselves. Nevertheless, the considerations presented above do not conclusively rule out future interpreters claiming so. But the burden of proof would seem to fall upon interpreters who might want to attribute this logically very demanding sense to the claim.⁵⁰

Notes

- 1 I will use the following abbreviations: *CPJ*: *Critique of the Power of Judgement*; *CPR*: *Critique of Pure Reason*; *GMS*: *Groundwork of the Metaphysics of Morals*; *MS*: *The Metaphysics of Morals*; *V-Th/Pöhlitz*: *Lectures on the Philosophical Doctrine of Religion*. In all cases I will quote from the translations of the Cambridge Edition of the Works of Immanuel Kant. Except in the case of the *Critique of Pure Reason*, where I will use the customary A/B pagination, references are to volume and page numbers in the Akademie edition.
- 2 The three formulas Kant refers to appear to be FLN, FH and FKE, which he goes on to discuss in the next passage and contrast with the ‘universal formula of the categorical imperative’ (Kant 1996a: 86; *GMS*, 4: 436), presumably FUL. But FLN and FKE are introduced as variants of their respective canonical formulations, FUL and FA. So, except in the titles of sections or where reference is in fact to a specific variant, I will speak for the sake of brevity about FUL and FA to refer to both the canonical formulation and to its variant. For discussion of whether FUL is indeed the ‘universal formula’ see Reich 2001: 135–6; Wood 1999: 163–5, 187–90; Geismann 2002: 377–8; Wood 2008: 82–4; Geiger 2010: 285–6, n. 22. For the claim that FA is not a formula of the categorical imperative see Geismann 2002: 381–3.
- 3 The translation of the above quote from *GMS*, 4: 436.9–10 is Gregor’s (Kant 1996a). In other translations, the phrase is taken to refer to any one of the three formulas by Watson, Abbott, Beck, Gregor and Zweig and to one formula by Friedrich, Paton, Ellington, Wood, Engstrom and Timmermann.
- 4 Ideally, it seems to me, a translation should preserve the ambiguity. Allison claims that Paton and Ellington do this. See Allison 2011: 246, n. 14. I do not see the ambiguity in their translations.
- 5 See Sedgwick 2008: 133.
- 6 In sections 1.1, 1.3 and 1.4, I use and develop arguments presented in Geiger 2010. My focus here is on the implications of O’Neill’s view for the question of the relation between the moral guidance of FUL and FH.
- 7 Atwell and O’Neill e.g. focus on this part of the sentence and read it as asserting the equivalence of the formulas in this sense. Atwell 1969: 273–4; O’Neill 1989c: 127.
- 8 It is not uncommon for readers to claim that an example is convincing or more plausible under one of the formulas only. In contrast, Ebbinghaus (1988: 209–19) explains at length that FLN and FH are equivalent in the sense that Kant’s four examples can be derived from both. But the point of this section is precisely to show that, on its prevailing reading, there is no *general* equivalence between them.
- 9 Kant distinguishes contradictions in thought or conception from contradictions in the will (Kant 1996a: 75; *GMS*, 4: 424). For the arguments I put forward in this section the former alone are relevant.
- 10 For discussion and some pertinent examples see e.g. Herman 1993b: 113–31; Herman 1993c: 138–43; Korsgaard 1996a: 82–5, 89–90, 97–101; Wood 1999: 102–7; Illies 2007: 313–14, 315; Allison 2011: 190–203.
- 11 O’Neill 2004: 96. See also: O’Neill 1991: 176–7; Illies 2007: 311–12. Korsgaard claims that employing a conception of rational willing that has moral content would undermine Kant’s project of establishing the connection between morality and rationality (1996a: 79–80). Kitcher argues that maxims are tested by FUL without their motivational component, which determines their moral worth (2004: 565–6). Timmons reviews different interpretations of FUL. None holds that it presupposes morally normative terms (2006: 194–6).
- 12 O’Neill 1989a: 84. See also O’Neill 1989b: 112.

- 13 See Herman 1993b: 116.
- 14 O'Neill 1989a: 84.
- 15 O'Neill 1989a: 81. See also O'Neill 1989c: 131–2.
- 16 O'Neill 1989a: 96.
- 17 O'Neill 1989c: 132–3.
- 18 O'Neill 1989c: 133.
- 19 See O'Neill 1989c: 138–9 and 1989b; Korsgaard 1996a: 92, 100–1; 1996b: 127.
- 20 See O'Neill 1989c: 138–40.
- 21 O'Neill 1989b: 109.
- 22 O'Neill 1989b: 109.
- 23 O'Neill 1989b: 109. See also Korsgaard 2006d: 309, n. 42.
- 24 O'Neill 1989b: 105.
- 25 See Wood 1999: 100; Allison 2011: 179.
- 26 For discussion of the precise sense Kant gives to the term 'derivation' see Krausser 1968.
- 27 O'Neill 1989c: 134. See also Nell (O'Neill) 1975: 7–8; Ebert 1976: 578.
- 28 O'Neill 1989a: 84. See also: O'Neill 1989c: 128; Krausser 1968: 326. Cf. Engstrom 2009: 220–3.
- 29 Herman 1993a: 77. See also Höffe 1977: 356, 359, 360; Nisters 1993: 90–1.
- 30 See Rawls 2000: 164–75; Rohlf 2009: 355–7.
- 31 See Wood 1999: 76–155. Wood develops further this approach in 'Universal Law' (unpublished manuscript).
- 32 Engstrom 2009: 172–8.
- 33 Engstrom 2009: 212.
- 34 Engstrom 2009: 214.
- 35 Engstrom 2009: 220.
- 36 Geiger 2010: 280.
- 37 For criticism of this paper and my response to it see Sticker 2015; Geiger 2015.
- 38 See O'Neill 1989c: 128; Wood 1999: 187. An elegant way to do this is to claim that our maxims qualify as universalizable only if they have rational nature as their end. See Sedgwick 2008: 149 (119–20, 128).
- 39 See: Engstrom 2009: 151; Allison 2011: 247. Cf. O'Neill 1989c: 131.
- 40 In the rare case where he does use it, he is thinking of the unity of a single end of all agents and not of the systematic unity of all the ends of all agents: 'if all human beings speak the truth, then among them a system of ends is possible; but if only one should lie, then his end is no longer in connection with the others' (Kant 1996b: 428; V-Th/Pölitz, 28.2: 1100). This brings FA very close to FUL. See Wood 1999: 167–8, cf. 186.
- 41 Cf. Wood 1999: 168–9.
- 42 I discuss this as well as the previous passage from the *Metaphysics of Morals* in the context of Kant's theory of moral feelings in Geiger 2011: 301–3.
- 43 See Korsgaard 1996c: 190–4.
- 44 See Wood 1999: 91.
- 45 See Ebbinghaus 1988: 219–20.
- 46 I take it that it is the analogy between FLN, FH and FKE and the categories of quantity that brings the abstract idea of morality closer to intuition and feeling (Kant 1996: 85; GMS, 4: 436). Cf. Reich 2001: 136–46. That the three formulas do offer substantially different perspectives on our duties is strongly suggested by the claim that to bring a duty closer to intuition it is best to consider it from all three perspectives.
- 47 See Wood 1999: 185–6; Sedgwick 2008: 150, n. 48; Allison 2011: 250.
- 48 I thank Allen Wood for this point.

- 49 This claim conflicts with the earlier statement: ‘Now, what serves the will as the objective ground of its self-determination is an end’ (Kant 1996a: 78; *GMS*, 4: 427). But Kant must mean ‘subjective ground’. See Paton 1964: 138, n. 2.
- 50 Uri Eran and Allen Wood commented extensively on drafts of this article and I am very deeply indebted to their insights. I earlier presented much shorter versions of it at the 2nd Biennial Meeting of the North American Kant Society at Cornell University and at the annual meeting of the Israel Philosophy Association at Tel-Aviv University. I thank Marilia Espirito Santo for her insightful comments read at the Cornell meeting. I am also very grateful for the helpful questions asked on these two occasions by Stephen Engstrom, Yakir Levin, Sally Sedgwick and Rachel Zuckert. Finally, I would like to thank two anonymous readers for their very helpful responses to the article.

References

- Allison, Henry E. (2011) *Kant's 'Groundwork for the Metaphysics of Morals': A Commentary*. Oxford: Oxford University Press.
- Atwell, John E. (1969) ‘Are Kant’s First Two Moral Principles Equivalent?’. *Journal of the History of Philosophy*, 7, 273–84.
- Ebbinghaus, Julius (1988) ‘Die Formeln des kategorischen Imperativs und die Ableitung inhaltlich bestimmter Pflichten’. In *Gesammelte Schriften*, vol. 2, *Praktische Philosophie 1955–1972* (Bonn: Bouvier Verlag), pp. 209–29.
- Ebert, Theodor (1976) ‘Kants kategorischer Imperativ und die Kriterien gebotener, verbotener und freigestellter Handlungen’. *Kant-Studien*, 67, 570–83.
- Engstrom, Stephen (2009) *The Form of Practical Knowledge: A Study of the Categorical Imperative*. Cambridge, MA: Harvard University Press.
- Geiger, Ido (2010) ‘What is the Use of the Universal Law Formula of the Categorical Imperative?’. *British Journal for the History of Philosophy*, 18, 271–95.
- (2011) ‘Rational Feelings and Moral Agency’. *Kantian Review*, 16, 283–308.
- (2015) ‘How Do We Acquire Moral Knowledge? Is Knowing Our Duty Ever Passive? – Two Questions for Martin Sticker’. *British Journal for the History of Philosophy*, 23.
- Geismann, Georg (2002) ‘Die Formeln des kategorischen Imperativs nach H. J. Paton, N.N., Klaus Reich und Julius Ebbinghaus’. *Kant-Studien*, 93, 374–84.
- Herman, Barbara (1993a) ‘The Practice of Moral Judgment’. In *The Practice of Moral Judgment* (Cambridge: Cambridge University Press), pp. 73–93.
- (1993b) ‘Murder and Mayhem’. In *The Practice of Moral Judgment* (Cambridge: Cambridge University Press), pp. 113–31.
- (1993c) ‘Moral Deliberation and the Derivation of Duties’. In *The Practice of Moral Judgment* (Cambridge: Cambridge University Press), pp. 132–58.
- Höffe, Otfried (1977) ‘Kants kategorischer Imperativ als Kriterium des Sittlichen’. *Zeitschrift für philosophische Forschung*, 31, 354–84.
- Illies, Christian F. R. (2007) ‘Orientierung durch Universalisierung: Der kategorische Imperativ als Test für die Moralität von Maximen’. *Kant-Studien*, 98, 306–28.
- Kant, Immanuel (1996a) *Groundwork of the Metaphysics of Morals* and *The Metaphysics of Morals*. In Mary Gregor (trans. and ed.), *Practical Philosophy* (Cambridge: Cambridge University Press),
- (1996b) *Lectures on the Philosophical Doctrine of Religion*. Trans. Allen W. Wood. In Allen W. Wood and George di Giovanni (trans. and eds.), *Religion and Rational Theology* (Cambridge: Cambridge University Press).
- (1997) *Critique of Pure Reason*. Trans. and ed. Paul Guyer and Allen W. Wood (Cambridge: Cambridge University Press).

- (2000) *Critique of Power of Judgment*. Ed. Paul Guyer, trans. Paul Guyer and Eric Matthews (Cambridge: Cambridge University Press).
- Kitcher, Patricia (2004) 'Kant's Argument for the Categorical Imperative'. *Noûs*, 38, 555–84.
- Korsgaard, Christine M (1996a) 'Kant's Formula of Universal Law'. In *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press), pp. 77–105.
- (1996b) 'Kant's Formula of Humanity'. In *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press), pp. 106–32.
- (1996c) 'Creating the Kingdom of Ends: Reciprocity and Responsibility in Personal Relations'. In *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press), pp. 188–211.
- (1996d) 'The Reasons We Can Share: An Attack on the Distinction between Agent-Relative and Agent-Neutral Values'. In *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press), pp. 275–310.
- Krausser, Peter (1968) 'Über eine Unvermerkte Doppelrolle des Kategorischen Imperativs in Kants *Grundlegung zur Metaphysik der Sitten*'. *Kant-Studien*, 59, 318–32.
- Nell (O'Neill), Onora (1975) *Acting on Principle: An Essay on Kantian Ethics*. New York: Columbia University Press.
- Nisters, Thomas (1993) *Kants kategorischer Imperativ als Leitfaden humaner Praxis*. Freiburg: Verlag Karl Alber.
- O'Neill, Onora (1989a) 'Consistency in Action'. In *Constructions of Reason: Explorations of Kant's Practical Philosophy* (Cambridge: Cambridge University Press), pp. 81–104.
- (1989b) 'Between Consenting Adults'. In *Constructions of Reason: Explorations of Kant's Practical Philosophy* (Cambridge: Cambridge University Press), pp. 105–25.
- (1989c) 'Universal Laws and Ends-In-Themselves'. In *Constructions of Reason: Explorations of Kant's Practical Philosophy* (Cambridge: Cambridge University Press), pp. 126–44.
- (1991) 'Kantian Ethics'. In Peter Singer (ed.), *A Companion to Ethics* (Oxford: Blackwell Publishers), pp. 175–85.
- (2004) 'Kant: Rationality as Practical Reason'. In Alfred R. Mele and Piers Rawling (eds.), *The Oxford Handbook of Rationality* (Oxford: Oxford University Press), pp. 93–109.
- Paton, H. J. (1964) *Groundwork of the Metaphysics of Morals*. New York: Harper Torchbooks.
- Rawls, John (2000) *Lectures on the History of Moral Philosophy*. Cambridge, MA: Harvard University Press.
- Reich, Klaus (2001) 'Kant und die Ethik der Griechen'. In Manfred Baum, Udo Rameil, Klaus Reisinger and Gertrud Scholz (eds.), *Gesammelte Schriften* (Hamburg: Felix Meiner Verlag), pp. 113–46.
- Rohlf, Michael (2009) 'Kant on Determining One's Duties: A Middle Course between Rawls and Herman'. *Kant-Studien*, 100, 346–68.
- Sedgwick, Sally (2008) *Kant's 'Groundwork of the Metaphysics of Morals': An Introduction*. Cambridge: Cambridge University Press.
- Sticker, Martin (2015) 'The Moral-Psychology of the Common Agent - A Reply to Ido Geiger'. *British Journal for the History of Philosophy*, 23 (forthcoming).
- Timmons, Mark (2006) 'The Categorical Imperative and Universalizability (GMS II 421–424)'. In Christoph Horn and Dieter Schönecker (eds.), *Groundwork for the Metaphysics of Morals* (Berlin: Walter de Gruyter), pp. 158–99.
- Wood, Allen W (1999) *Kant's Ethical Thought*. Cambridge: Cambridge University Press.
- (2008) *Kantian Ethics*. Cambridge: Cambridge University Press.