

The Problem with Using a Maxim Permissibility Test to Derive Obligations

Introduction

Kant introduces his famous universalization formulations of the Categorical Imperative in the *Groundwork to the Metaphysics of Morals*:

Act only according to that maxim through which you can at the same time will that it become a universal law.

So act as if the maxim of your action were to become by your will a universal law of nature.¹

The purpose of this paper is to show that, if the universalization formulations are our only standards for judging right from wrong and permissible from impermissible, then we have no obligations. I shall do this by examining five different views of how obligations can be derived from the universalization formulations and arguing that each one fails. I shall argue that the first view rests on a misunderstanding of the universalization formulations; the second on a misunderstanding of the concept of an obligation; the third on a misunderstanding of the concept of a maxim; the fourth on a misunderstanding of the limits of action description; and the fifth on a misunderstanding of the universalization formulations again.

Two points of clarification. First, it might be thought that, if I am merely taking on different views of how obligations can be derived from the universalization formulations piecemeal, then, even if my criticisms of these views withstand critical scrutiny, my thesis will remain unproven, for there always could be another view out there, just of out reach of the ones I have considered. But this thought is mistaken, and I shall explain why when discussing view 4.

Second, it might be thought that, if my thesis is correct, then, although we never shall be able to figure out, using only the universalization formulations, what our obligations are, the universalizability tests can, in principle, demarcate them. The idea here might be that there is a difference between proof and truth, and my arguments impugn only the former. Perhaps infinitely many maxims would have to be tested in order to *prove* that one is obligatory and we, with our finite capacities, are not up to the task, or perhaps there would have to be a fixed enumeration of maxims in order to *derive* an obligation and there is no such enumeration. But, the thought continues, the universalizability tests demarcate our obligations all the same—the moral landscape might be mapped even if the map lies out of our reach.

However, this second thought is also mistaken. My thesis is not about proofs, deductions, or derivations *per se*. My thesis is about the actual moral landscape that we face, with or without a map, if the universalization formulations are our only tools. If we want to carve moral nature at the joints, and if any such project requires that there be a nonempty set of obligatory maxims, then the universalization formulations had better not be our only carver. Or so I aim to show, anyway.

Section 1: If a Maxim is Universalizable, then it is Obligatory

There is a reading of Kant's *Religion within the Boundaries of Mere Reason*, one that goes back at least to Fichte, according to which everything we do is either impermissible or obligatory. This is sometimes referred to as rigorism, and the crux of the reading that is relevant for present purposes is this:

VIEW₁ If a maxim is universalizable, then it is obligatory.

¹ GMS, AA 04: 421, emphases omitted. All citations to Kant are in accordance with the Prussian Academy pagination that runs in the margins of most translations. All translations are taken from the Cambridge Edition of the Works of Immanuel Kant.

According to VIEW_I, the universalization formulations do demarcate obligatory maxims from non-obligatory maxims. A maxim that fails the universalization formulations, like a maxim to commit suicide from self-love, or a maxim to make a lying promise in order to get some ready money, is impermissible. But a maxim that passes the universalization formulations, like a maxim to promote my natural talents, or a maxim to promote the happiness of all, is obligatory.

I find rigorism counterintuitive and, indeed, although Fichte seems to have found it attractive, those who ascribe it to Kant generally seem to do so as a preface to dismissing his ethics as absurdly misguided. And for precisely this reason, it will not do to reply to VIEW_I that, although, in principle, it undermines my thesis, it does so at the cost of the plausibility of the universalization formulations. Those who favor this reading of Kant only would double down in the face of such a response: "that," they might say, "is the point." Moreover, despite the fact that I think that any view that eliminates the merely permissible is dismissible, my thesis is not: either the universalization formulations cannot demarcate obligations or they should be discarded. My thesis is that the universalization formulations cannot demarcate obligations full stop. So, if my reply to rigorism were merely that it saddles Kant in general and the universalization formulations in particular with an unsustainable ethical outlook, that would not rebut VIEW_I among those with whom it is popular, nor would it help to prove my thesis.

Similarly, it will not do to point out that Kant clearly thinks that some things are morally indifferent and, thus, that Kant rejects rigorism. Kant clearly does think this, as may be seen from the following passage from the *Metaphysics of Morals*:

...the human being can be called fantastically virtuous who allows nothing to be morally indifferent...and strews all his steps with duties, as with mantraps; it is not indifferent to him whether I eat meat or fish, drink beer or wine, supposing that both agree with me. Fantastic virtue is a concern with petty details" which, were it admitted into the doctrine of virtue, would turn the government of virtue into tyranny.²

But this does not help because Kant might have misunderstood the universalization formulations. Kant was fallible, after all, and so the fact that he thought that there are merely permissible actions, conjoined with the fact that he formulated and subscribed to the universalization formulations, does not entail (although it perhaps does suggest) that VIEW_I is mistaken.

But VIEW_I *is* mistaken, and fortunately it is easy to see why. Not only does VIEW_I not comport with Kant's general remarks about ethics (or, for that matter, his general remarks about the universalization formulations), it seems to rest on an obvious misreading of the universalization formulations. The universalization formulations tell us to act *only* on maxims that are universalizable, or to act *as if* our maxims were universalizable. They do *not* tell us to act on *all* maxims that are universalizable, and only the latter can get us to VIEW_I. Thus, VIEW_I rests on a misunderstanding of the universalization formulations.

Section 2: If a Maxim of the Form 'I will Q' is Universalizable, then it is Obligatory to Q

From the previous section of this paper it may be seen that the universalization formulations seem to give the following biconditional:

A maxim M is permissible if but only if it is universalizable.

VIEW_I is consistent with this biconditional, for it also might be the case, depending on how we understand moral dilemmas, that if X is obligatory, then X is permissible. The problem with VIEW_I is that it misunderstands the universalization formulations in such a way as to eliminate the merely permissible. But from this it may be seen

² MS, AA 06: 409, emphases omitted.

that, if the universalization formulations are going to demarcate the obligatory, this cannot be solely on the basis of universalizability; it must be on the basis of universalizability+Y, where Y is some other mark.

Patricia Kitcher suggests a strategy for getting around this using universalizability+the form of a maxim. According to Kitcher, maxims can conform to either of the following schemata:

1. I will P, if I like.
2. I will Q.

For example, I might adopt the maxim to help my daughter with her homework if I like (schema 1), or I might adopt the maxim to keep my promises (schema 2). The point Kitcher wants to make is that, if an agent adopts a maxim conforming to schema 1, she might not P, whereas if an agent adopts a maxim conforming to schema 2, then she definitely will Q (*ceteris paribus*). This suggests that if a maxim that conforms to schema 1 is universalizable, then there is a permissive law—an agent may P or not depending on whether she likes—whereas if a maxim that conforms to schema 2 is universalizable, then there is an obligation—an agent ought to Q regardless of whether she likes. For present purposes, the important part of this may be summed up as a conditional:

VIEW₂ If a maxim of the form 'I will Q' is permissible, then Q is obligatory.³

To see the problem with this, suppose that the maxim 'I will Q' is universalizable and, thus (from the biconditional at the beginning of this section), permissible. Then (from VIEW₂) it follows that Q is obligatory. But the maxim might be merely permissible, whence it would follow that there are other permissible maxims that the agent may adopt. Suppose that R is such a maxim. But if the agent adopts R, then the agent might not Q and, crucially, her omission of Q would be permissible. Thus, whereas VIEW₂ entails that Q is obligatory, an agent nonetheless permissibly can omit Q, and that is no obligation at all. To put this another way, Q is going to be genuinely obligatory only if 'I will Q' is obligatory, and the latter is precisely what needs to be shown; even if all obligations are permissible, not all permissions are obligatory—a merely permissible obligation is an oxymoron. So, VIEW₂ seems to rest on a confusion about the concept of an obligation.⁴

Section 3: If it is Impermissible to Adopt Maxim M, then it is Obligatory to Adopt Maxim not-M.

In the previous section I argued against using permissibility to demarcate obligations. But our toolbox is not so sparse. The biconditional from the previous section (M is permissible if but only if it is universalizable) also gives us impermissibility. Moreover, the following biconditional seems relatively unobjectionable:

Maxim M is impermissible if but only if it is obligatory not to adopt M.

³ Here is how Kitcher puts it:

When considering laws of obligation, the 'willing' side of the argument for FUL is developed by considering the motive force of imperatives. An agent can derive the maxim, 'I to do ACP' only if he adopts (however temporarily) the law, 'Everyone ought to do ACP'... Parallel considerations enable us to continue the laws of permission side of the argument, although the case is more controversial. An agent can derive his maxim, 'I to do ACP, if I like' from a law of permission, 'Everyone to do ACP, if he likes' only if he adopts that law of permission. (Kitcher, 2004, 576)

⁴ Suppose that Kitcher changes her view (or, more charitably, that I have misunderstood her). Suppose Kitcher drops the action/maxim distinction and stipulates that, if a maxim of the form 'I will Q' is universalizable, then that *maxim* is obligatory. Call this VIEW₂*. VIEW₂* gets around the objection in the paragraph to which this note is appended. But it still does not work. Quite apart from the usual problems associated with stipulation, VIEW₂* will fall prey either to the problems associated with VIEW₃ or those associated with VIEW₄, depending on how we further specify the details. We also might question whether, in introducing such a stipulation, VIEW₂* is really working with the universalization formulations anymore or whether it has moved onto something inspired by the universalization formulations.

This suggests something like De Morgan's law for maxims, where we drive the negation in from the deontic operator to the action in a maxim. That is, we can start with the fact that it is impermissible to adopt some maxim M, where M has the form 'I will Q,' and we can conclude from this that it is obligatory to adopt the maxim 'I will not-Q.' If Q is 'never to help anyone,' then, because this maxim is not universalizable and, thus, impermissible, we can drive the negation in to get an obligation to adopt the maxim 'sometimes help someone,' and thereby we get an obligation to help others. This seems to be what John Harrison has in mind when explaining how the universalization formulations generate obligations:

Kant holds (a) that a maxim is not morally acceptable and must not be adopted...if it cannot be universalized; (b) that it may be adopted...if it can be universalized. He may also have thought (c) that a maxim must be adopted (ought to be adopted) if what I shall loosely call its 'contradictory' is not universalizable... (c) would follow from (a) together with the additional premise, which I see no reason to cavil about, that if it is our duty not to do A, then it is our duty to do non-A.⁵

It also seems to be what Barbara Herman has in mind when she explains why the maxim 'to help some others sometimes' is obligatory on the universalization formulations:

[I]f the CI procedure shows that it is impermissible to adopt the maxim, 'to never help anyone,' it follows that we must adopt its contradictory, 'to help some others sometimes.'⁶

We can summarize the core of this view in another conditional:

VIEW₃ If it is impermissible to adopt maxim M, then it is obligatory to adopt maxim not-M.

The details of VIEW₃ then can be filled in with the maxim schemata from the previous section of this paper. If a maxim of the form 'I will P, if I like' (schema 1) is impermissible, as is the maxim 'I will break my promises if it suits my purposes,' then its contradictory is obligatory. Using our De Morgan's law for maxims, we may conclude that 'I will keep my promises' is obligatory. Alternatively, if a maxim of the form 'I will Q' (schema 2) is impermissible, as is the maxim 'I will never help anyone,' then, as above, we may use our De Morgan's law for maxims to conclude, with Herman, that 'I will help some others sometimes' is obligatory.

To begin to understand where VIEW₃ goes wrong, note that it has false positives. This is not because the universalization formulations misclassify maxims. It will be noted that, despite the large literature on how to interpret the universalization formulations and, more relevantly for present purposes, whether any such interpretation can get around the myriad false positives and false negatives with which they have been charged, I have not said anything about this, nor do I intend to do so now. I do not need to. The reason VIEW₃ has false positives has nothing to do with the universalization formulations *per se* and everything to do with how we are proposing to derive obligations from them. For example, the maxim 'to drive into pedestrians' is impermissible, and from VIEW₃ it follows that it is obligatory to adopt some maxim about not driving into pedestrians. That might seem all well and good until we realize that, for all that has been said, this obligation holds for any agent, anywhere, at any time. This exposes VIEW₃ to *reductio*: cars are a relatively recent invention, and it would be absurd to assert that our neanderthal forebears were obligated to adopt any maxim whatever about them.

The most immediate problem with VIEW₃ is that obligations are agent-sensitive in a way that impermissibility are not. To use the example in the previous paragraph to illustrate, it is impermissible for any agent at any time and anywhere to adopt the maxim 'to drive into pedestrians,' but (as just noted) there is no corresponding obligation. The deeper problem with VIEW₃, the one that illuminates this immediate problem, has to do with the way that VIEW₃ understands maxims. In the conditional above, I used the term 'not-M' to echo Harrison, but we just as

⁵ (Harrison, 1957, 52).

⁶ (Herman, 1993, 63).

easily could have used Herman's 'maxim contradictory.' Herman's term is revelatory because it invokes the idea of propositional contradictories, where propositions P_1 and P_2 are contradictories if and only if (P_1 is true if and only if P_2 is false). Herman's thought might be that an agent adopts 'to help some others sometimes' if and only if she does not adopt 'never to help anyone,' and, thus, the impermissibility of the latter entails the obligatoriness of the former. The corresponding metaphysical view can be filled in using Leibnizean (not Kantian) possible worlds semantics. That is, if there are maximally consistent sets of sentences corresponding to each possible world, then presumably there are maximally consistent sets of sentences corresponding to each possible agent, much like Leibnizean completely determined concepts. Then, for any given agent and action, there is a fact of the matter, even if it is beyond our ken, whether and under what conditions the agent would perform that action, and from this it follows that there is a fact of the matter, again perhaps beyond our ken, for any agent and any maxim, including but not limited to neanderthals and driving-maxims, whether an agent has adopted it—and if an agent has not adopted the maxim M , then she will have adopted its contradictory.⁷

But maxims do not work like this. Agents need not consciously adopt, deliberate about, or even be aware of their maxims. But maxims are not completely external descriptions of an agent either. It might be true that none of us ever floats up off the floor like a soap bubble. But it is probably also true that none of us has a maxim to refrain from doing so. That is not because it would be impossible to adopt such a maxim. Recall O'Brien's assertion in 1984 that, if he wished, he could float in just this way. I do not want to debate the finer points of O'Brien's idealism—I merely want to note that it would be possible, although not advisable, to work your way into a mindset in which you end up adopting such a maxim. So, there are maxims such that we adopt neither them nor their negations. This is why obligations can be agent-sensitive in a way that impermissibilities are not, and it is why VIEW₃ does not work. From the fact that a given maxim is impermissible, exactly nothing follows about the obligatoriness of any other maxim because an agent can fulfill an impermissibility by doing exactly nothing. The truth of an external description about an agent (e.g., she is not stealing while she is sleeping) does not entail that the agent has adopted any corresponding maxim unless we misunderstand the concept of a maxim.⁸

Section 4: If Maxim M is Impermissible and Agent A is Deliberating About M, then it is Obligatory for A to Adopt Maxim Not-M

Although Herman's appeal to maxim contradictories seems to rest on a misunderstanding of the concept of a maxim (i.e., maxims, unlike propositions, do not have contradictories), there is a premise lurking in the shadows of VIEW₃ that might be salvaged:

If all but one of an agent's options are impermissible, then the one permissible option is obligatory.

But now, if an agent is deliberating about some course of action, then, it might be thought, this conditional can be used to derive an obligation without running afoul of concerns about the concept of a maxim. This seems to be what Korsgaard has in mind in the following passage:

Kant does derive obligatory ends from the Formula of Universal Law, but he does it by a curiously roundabout procedure in which someone is imagined formulating a maxim of rejecting them and then finding it to be impermissible. This argument does not show that there would be a moral failing if the agent merely unthinkingly neglected rather than rejected these ends. The

⁷ There is some oversimplification here. For example, the move from maximally consistent sets of sentences corresponding to possible worlds to maximally consistent sets of sentences corresponding to possible individuals in those worlds does not work on account of the fact that many (indeed, perhaps all) of these individuals will inhabit multiple possible worlds. This raises rather deep metaphysical problems for Leibniz's monadology, but none that needs to be discussed right now.

⁸ Indeed, not only might an agent adopt neither a maxim nor its putative contradictory, but, more, whether an agent *would* adopt a given maxim or its putative contradictory might be entirely indeterminate. This ties into the issues gestured at in note 7 above.

point about the pervasiveness of these ends in the moral life is a more complicated one, one that follows from their adoption by this route...⁹

We can summarize the core of this, building off VIEW₃ from the previous section, in the following conditional:

VIEW₄ If maxim M is impermissible and agent A is deliberating about M, then it is obligatory for A to adopt maxim not-M.

The problem now is that, for any given course of action about which an agent is deliberating, there will be infinitely many different options, many of which will involve more specific or more general descriptions of the action in question, and many of which will involve alternative actions altogether. Further, if there is one permissible option, then there always will be a plurality permissible options, and so there never will be an obligatory one. To see why, suppose that action A is permissible under description B. Then there must be some more specific description of A that is also permissible, whence it follows that there is more than one permissible option.

An example will bring this closer to intuition. Suppose that an agent is deliberating about whether to help at a soup kitchen. Here is a small sample of maxims she might adopt:

1. Never to help at soup kitchens.
2. Never to help at soup kitchens which serve meat.
3. Never to help at soup kitchens which are run inefficiently.
4. Never to help at soup kitchens where they make volunteers clean dishes.
5. Never to help at soup kitchens where they serve unhealthy food.
6. Sometimes to help at some soup kitchens.
7. Sometimes to help at local soup kitchens.
8. To help at local soup kitchens when their staffing is low.
9. To help at local soup kitchens on the weekends.
10. To help at local soup kitchens when at least one friend is able to volunteer with you.

I guess that quite a few of these (at least 6-10, and probably 2, 3, and 5) are permissible. But then it follows immediately that none of them is obligatory. And these 10 do not exhaust the options; obviously there are more.

This might push some people back to VIEW₃. The advantage of VIEW₃ is that this plurality of maxims is irrelevant. Either an agent will A or not (she will do non-A), and so if one of these (A or non-A) is permissible and the other is not, then the permissible one is obligatory. If we try to inject further details into the action description, that does not undermine the VIEW₃ line of reasoning; in fact, we can duplicate it: either an agent will <A+details> or not (she will do non-<A+details>), and so if one of these (<A+details> or non-<A+details>) is permissible and the other is not, then the permissible one is obligatory. The problem with this line of reasoning is not, as we saw, the fact that more than one of these action descriptions always will be permissible; rather, it is that these action-descriptions are purely external, not maxims. We can, to be sure, assess them for permissibility using the universalization formulations as if they were maxims. But precisely because they are not maxims or, rather, precisely because when we treat them as maxims, an agent need not adopt either one of them, and so no facts about obligations follow from such assessments, which is why VIEW₃ is a wash. So, of course, retreating from VIEW₄ to VIEW₃ does not help.

And now I can make good on a promissory note from the introduction to this paper. Thus far, I have been explaining why different attempts to derive obligations from the universalization formulations fail. But in the introduction, I said that I would explain more: why all attempts are foredoomed to fail—why there can be no obligations if the universalization formulations are our only standards for judging right from wrong and

⁹ (Korsgaard, 1996, 152-153).

permissible from impermissible. It is because, for any given permissible maxim, there is always another permissible one lurking around the corner, whence it follows that there are no obligatory maxims—and so any attempt to derive obligatory maxims will be unsuccessful.

Section 5: If Agent A Adopts End E, Then A Ought to Universalize E

Some might object at this point (indeed, some have objected at this point when I discuss these arguments with them) that ends, like the happiness of others or one's own perfection, are not like actions, like helping at a soup kitchen or keeping a promise, and there is a straightforward way to derive obligatory ends using the universalization formulations. Indeed, the derivation is inspired by Kant's derivation of the duty of beneficence in the *Metaphysics of Morals*, and (the objection continues) Kant does not think that actions or maxims are obligatory anyway—only ends are. So I want to discuss one last view before I wrap up this paper.

The argument that my interlocutors have in mind is something like this: suppose you adopt end E. For E to be permissible, you must universalize it. So, if you adopt end E, then you are obligated to universalize E. More concretely, if you have your own happiness as an end, then you are obligated to adopt the happiness of all as an end. As in the previous sections, we can summarize the core of this view as a conditional:

VIEW₅ If agent A adopts end E, then A ought to universalize E.

Now, strictly speaking, there is no duty to adopt the universalized E on VIEW₅. An agent could (permissibly) just give up the non-universalized end E. But if the agent is unable (or simply unwilling) to give up E, then the obligation follows. Kant thinks that agents are unable to give up the end of their own happiness, and he seems to think that happiness is the only end of this kind, which gives the duty of beneficence a special status in his system. It is for this reason (i.e., because happiness is the only end that agents cannot give up) that Allen Wood maintains that the duty of beneficence is the only obligation that can be derived from the universalization formulations.¹⁰

But VIEW₅ does not work any more than the previous ones. The first indication that something has gone wrong with VIEW₅ is that the reasoning is not generalizable. For example, suppose that I have the end of becoming a professor. It does not (or, at any rate, should not) follow from this that I ought to have the end of everyone becoming a professor (or give up the end myself). The second indication that something has gone wrong is that there are lots of other permissible ends that I might adopt *instead* of the end of the happiness of all, like the end of the happiness of my friends and family. This latter (i.e., the happiness of my friends and family) *must* be a permissible end, for I could adopt it as a permissible *subordinate* end, subordinate to the end of the happiness of all. So I could just adopt this end *without* the superordinate end of the happiness of all.¹¹

But this only indicates that something has gone wrong with VIEW₅. We still have to figure out what it is that has gone wrong. And I think it is in the step that says: if you have an end, then you must universalize that end. I do not think that this follows from the universalization formulations. I think that, if you have an end, you have to determine whether it is universalizable—and if it is not, then you ought to divest yourself of that end. But that is not the same as saying you must in some way universalize it—i.e., give it universal scope in the way that the above derivation does for the end of happiness. To be sure, adopting the happiness of all as an end might be *one way* of permissibly promoting my own happiness. But so is adopting the happiness of my friends and family as an end, and so is adopting the happiness of those within my sphere of influence—and so on. The universalization formulations

¹⁰ (Wood, 1999, chapter 3 section 7.2).

¹¹ There are two things that are notable about this point: (1) if we try to argue that this end (the happiness of friends and family) is permissible *only* if subordinate to the happiness of all, then we must be appealing to something *other* than the universalizability tests—we are no longer assessing the end (maxim of ends) by the universalizability tests, we are assessing the end *by its context*. Barbara Herman does this in some of her publications (i.e., she says that what makes a maxim impermissible is the maxim it is subordinate to), but, again, this is no longer appealing (solely) to the universalization tests (see Herman, 1993, 65, discussed briefly in Kahn, 2021, section 4). (2) Once we start thinking about this—that is, about the difference between the happiness of all and the happiness of friends and family—I think it becomes implausible to maintain that many (if any) actually have adopted the end of the happiness of all (rather than some more restricted end).

are applied to maxims of ends in the same way that they are applied to maxims of actions, and the same considerations that were relevant in assessing VIEW₄ are, thus, relevant in assessing VIEW₅.

Conclusion

In this paper, I have argued that if the universalization formulations are our only moral standards, then we have no obligations. I did so by assessing five different views of how obligations can be derived from the universalization formulations. I argued that the first rests on a misunderstanding of the universalization formulations; the second on a misunderstanding of the concept of an obligation; the third on a misunderstanding of the concept of a maxim; the fourth on a misunderstanding of the limits of action description; and the fifth on a misunderstanding of the universalization formulations again. Along the way, I tried to explain why it is not only these views that fail—I tried to explain that, once we understand the reasons for the reasons why these views fail, we can see that there are no obligations at all, at least if the universalization formulations are our only moral standards. However, it should be borne in mind that nothing I have said here has in any way undermined whether the universalization formulations can demarcate the permissible and the impermissible. And that might be all that we need them to do anyway.

Bibliography

Harrison, Jonathan. 1957. "Kant's Examples of the First Formulation of the Categorical Imperative." *Philosophical Quarterly* Vol. 7, No. 26: 50-62.

Herman, Barbara. 1993. *The Practice of Moral Judgment*. Harvard University Press.

Kahn, Samuel. 2021. "Obligatory Actions, Obligatory Maxims." *Kantian Review* Vol. 26, No. 1: 1-25.

Kitcher, Patricia. 2004. "Kant's Argument for the Categorical Imperative." *Nous* Vol. 38, No. 4: 555-584.

Korsgaard, Christine. 1996. *Creating the Kingdom of Ends*. Cambridge University Press.

Wood, Allen. 1999. *Kant's Ethical Thought*. Cambridge University Press.