

Positive Duties, Kant's Universalizability Tests, and Contradictions

Samuel Kahn

Indiana University-Purdue University, Indianapolis

In this paper I am going to raise a problem for recent attempts to derive positive duties from Kant's universalizability tests. In particular, I argue that these recent attempts are subject to *reductio* and that the most obvious way of patching them renders them impracticable. I begin by explaining the motivation for these attempts. Then I describe how they work and begin my attack. I conclude by considering some patches.

I. The Problem

When I speak of Kant's universalizability tests, I am referring to two formulations of the Categorical Imperative. One of these tests says that agents ought to act only on maxims that they can will at the same time as universal laws. The other says that agents ought to act only on maxims that they can will at the same time as laws of nature.

There are important differences between these two tests, but for present purposes these differences may be overlooked. I shall say that a maxim is universalizable if but only if it passes these two tests, and I simply shall assume that the two tests are extensionally equivalent. For present purposes what is important is that these tests seem to be tests of permissibility. To put the point schematically, these tests yield the following two biconditionals: (1) a maxim is permissi-

ble if and only if it is universalizable; and (2) a maxim is impermissible if and only if it is not universalizable.

If this is correct, then even if the universalizability tests have neither false positives nor false negatives, they yield only two deontic categories: permissible and impermissible. This is a problem, however, because pretheoretic intuition indicates that there are more such categories, from the supererogatory to the suberogatory. Indeed, Kant's own ethics is often characterized as an ethics of *duty*. It is therefore somewhat surprising to note that the universalizability tests do not have a separate mark for the category of the obligatory. To borrow a turn of phrase, the universalizability tests are not expressively powerful enough to carve morality at the joints: some joints can be carved, but too many are missing. It is like trying to do exponentiation using Robinson arithmetic; in a best case scenario, the universalizability tests require supplementation.

II. A Recent Approach

Exactly this kind of supplementation is supplied by Onora O'Neill. Her proposal may be separated into three steps. First, O'Neill introduces the notion of a maxim contrary.

A maxim is a subjective principle of volition, like "I will to make a lying promise when in financial distress to get some ready money," and it is not propositional in nature. So O'Neill defines maxim contraries in terms of universalized typified maxim counterparts (UTCs).

The idea is that, for any maxim, we can imagine that maxim as having universal scope (instead of being merely subjective) and as a description of what actually happens (instead of

being merely volitional). For example, the UTC of the maxim above would be: “Everybody will make a lying promise when in financial distress to get some ready money.”

Because a UTC is propositional, the traditional concept of a contrary may be used to define maxim contraries. That is, two UTCs are contraries if but only if they cannot both be true, and O’Neill says “I shall call the pair of maxims which must be tested...contraries, because their UTC’s are contraries” (O’Neill, 2013, 162n25).

The second step of O’Neill’s approach, implicit in the quotation just reproduced, is to apply the universalizability tests to maxim-contrary *pairs* rather than merely to a maxim *simpliciter*. Whereas previously there were two categories, either a maxim is universalizable or it is not, now there are three: (1) one of the two, the maxim or its contrary, is universalizable, and one of the two is not; (2) a maxim and its contrary are both universalizable; and (3) neither a maxim nor its contrary is universalizable.

The third and final step of O’Neill’s approach is to propose that these three categories can be used to capture our intuitive categories of the obligatory, the impermissible, and the merely permissible. O’Neill maintains that category (1) yields obligatory and impermissible: the universalizable one of the two is obligatory whereas the nonuniversalizable of the two is impermissible. And O’Neill maintains that the maxims in categories (2) and (3) are merely permissible.

To see how this works, consider the following six maxims:

- A. I will to eat small children alive.
- B. I will not to eat small children alive.
- C. I will to play solitaire when I feel the urge.

D. I will not to play solitaire when I feel the urge.

E. I will to buy toy trains and never to sell them.

F. I will to sell toy trains and never to buy them.

Now permit me to make some assertions for illustrative purposes. A, E, and F are not universalizable. B, C, and D are universalizable. But A and B are contraries, B and C are contraries, and E and F are contraries.

From this it follows that A is impermissible, B is obligatory, and C-F are merely permissible. These categorizations seem to agree with pretheoretic intuition, and so the three-step theory is confirmed.

Maxims E and F are unusual, and O'Neill has a special name for them: she calls them non-reciprocal action maxims. She admits that assigning maxims in category 3 to be merely permissible is an extrapolation from Kant: Kant would not have accepted that a maxim can fail the universalizability tests and nonetheless be permissible. But O'Neill justifies this extrapolation on the grounds that it is necessary to avoid counterintuitive results, like assigning E and F to be impermissible (O'Neill, 2013, p. 165).ⁱ

I think that the book in which O'Neill advances this proposal has been so influential that engaging it would be worthwhile even if nobody had followed her lead. But as a matter of fact, many Kantians and Kant scholars do seem to subscribe to something like O'Neill's approach, at least with regard to categories 1 and 2. For example, Hernandez asserts that in Kantian ethics an action is obligatory if but only if its "contrary maxim conflicts with the moral law" (Hernandez, 2010, p. 403); Freyenhagen contends that "[a] strict duty is one where universalizing a maxim

contrary to it would generate a contradiction in conception, while a wide duty is one where universalizing a maxim contrary to it would avoid this problem, but instead generate a contradiction in willing” (Freyenhagen, 2012, p. 56); and Guyer maintains that “if I must reject the maxims of letting all my talents rust or never helping anyone else, then I must accept their logical contraries, namely, maxims of cultivating at least some of my talents and helping at least some other people some of the time” (Guyer, 2006, p. 194).

III. My Attack

To see how a problem arises now, consider the following four maxims:

- G. I will to supplement my vegan diet with the occasional human child.
- H. I will to supplement my vegan diet with the occasional human adult.
- I. I will to supplement my vegan diet with the occasional mealworm.
- J. I will to supplement my vegan diet with the occasional termite.

Notice that all of these maxims are pairwise contraries. That is, G and H are contraries, and so are G and I and G and J, etc. To see this, construct the corresponding UTCs: no two UTCs from these maxims can be true at the same time.

Now I assert that G and H are not universalizable whereas I and J are. I am not going to defend this assertion by advancing a detailed interpretation of the universalizability tests.ⁱⁱ What I want to say is that if this assertion is false, then the universalizability tests have false positives

and false negatives. That would call into question the rationale for engaging in a supplementation project like O'Neill's to begin with. So denying this assertion would create more problems than it would solve.

However, if this assertion is true, then O'Neill's proposal is subject to *reductio*. To see why, note the following three facts:

1. Maxims G and H, by virtue of being pairwise contraries that are both nonuniversalizable, fall into category 3 and therefore are both merely permissible.
2. Maxims I and J, by virtue of being pairwise contraries that are both universalizable, fall into category 2 and therefore are both merely permissible.
3. Maxims G and I are pairwise contraries and G is not universalizable whereas I is universalizable, whence it follows that they fall into category 1 and G is impermissible and I is obligatory.

From 1 and 3 it follows that maxim G is merely permissible and impermissible. From 2 and 3 it follows that maxim I is merely permissible and obligatory. Similar results follow regarding maxims H and J. But no maxim can be both merely permissible and impermissible or merely permissible and obligatory. Can this be patched?

IV. Some Patches

I want to be very clear about the problem that I am pointing to in section III. The problem is *not* that maxims G and H are permissible. That *is* a problem. But that is *not* the problem I am pointing to. The problem I am pointing to is one of internal incoherence: the deontic categories <merely permissible>, <impermissible>, and <obligatory> are mutually exclusive.

Nonetheless, the fact that G and H are permissible on this account suggests a way to kill two birds with one stone. If O'Neill's proposal is modified so that in at least some cases when a maxim and its contrary both fail the universalizability tests they are impermissible, as with G and H, that would avoid the contradiction associated with assigning G and H to be both merely permissible and impermissible.

But that would not eliminate all of the contradictions derived above. In addition to maxims G and H being assigned to <merely permissible> and <impermissible>, maxims I and J also were assigned to mutually exclusive deontic categories: maxims I and J were both merely permissible and obligatory.

One easy patch, along the lines of the one just suggested, would be to modify O'Neill's proposal so that when a maxim and its contrary are both universalizable, they are permissible rather than merely permissible. The categories of the obligatory and the merely permissible are generally taken to be species of the genus <permissible>. So the obligatory and the permissible are not mutually exclusive even though the obligatory and the *merely* permissible are. So if O'Neill's proposal is modified in this way, the contradictions alluded to in the previous paragraph are dissolved.

Now some might object on moral grounds to the claim that maxims I and J are obligatory. This objection, as I am thinking of it, is to the idea that morality would make it obligatory to eat

a vegan diet supplemented only with the occasional mealworm or only with the occasional termite. That is, it might be argued that a plausible moral principle will give the result that it is merely permissible to eat a vegan diet supplemented with the occasional mealworm; that it is merely permissible to eat a vegan diet supplemented with the occasional termite; that it is merely permissible to eat a vegan diet supplemented with the occasional mealworm or the occasional termite; and so on—that it is merely permissible to have any sort of diet at all, from fruitarian to raw vegan to whatever, provided the food one is eating is ethically sourced.

But that is not the objection I want to make. The objection I want to make does not depend on intuitions about the moral status of eating (or other) practices: the objection I want to make is that even if the two patches suggested in the previous paragraphs are incorporated into O'Neill's proposal, it is still subject to *reductio*.

For this final *reductio* I need the following principle:

If it is obligatory to A and if it is impossible to A and to B, then it is impermissible to B.ⁱⁱⁱ

I cannot defend this principle at length here. I note only that (a) I find this principle intuitively plausible; and (b) this principle is implied by a thesis that seems to be widely accepted, especially among Kantians: that it is obligatory to take the necessary means to satisfy an obligation.^{iv} But if this principle is granted, then the fact that I is obligatory entails that J is impermissible and *vice versa*, whence it follows that I and J are both obligatory and impermissible.

One might continue to patch O'Neill's proposal along the lines of the two patches already offered. For example, perhaps universalizable maxims in category 1 should not be considered

obligatory. But this risks undermining the original purpose of the proposal: to enrich the expressive power of the universalizability tests so that more than two deontic categories can be enunciated. Moreover, this continued patchwork also risks being *ad hoc*. A more useful way forward, I think, is to diagnose what is giving rise to these problems.

In fact I think the root of it all is relatively easy to see at this point. There are two sources. First, any given maxim is going to have more than one contrary. Second, categories 1 and 2 and categories 1 and 3 can overlap. To make this clear, suppose there is a set of maxim contraries *M*. Now suppose that (i) more than one maxim in that set is universalizable and (ii) more than one maxim in that set is not universalizable. Then contradictions will be derivable in just the way they were derived in this paper.

Fortunately, this suggests a final patch to O'Neill's proposal, one that, I think, would not be *ad hoc* and that would preserve, at least in principle, the category of the obligatory: instead of saying that if a maxim is universalizable and its contrary is not then the maxim is obligatory, it might be asserted that if a maxim is universalizable and *none* of its contraries is universalizable then the maxim is obligatory.^v

However, this patch preserves the category of the obligatory *only* in principle. Given our finite cognitive capacities, no human ever is going to be in a position to test all of the contraries of a maxim. So no human ever is going to be in a position to determine that a maxim is obligatory using this proposal. And because that is exactly what this proposal was supposed to enable, this is a pretty serious problem.

Bibliography

Freyenhagen, Fabian. (2012) "The Empty Formalism Objection Revisited." In Thom Brooks (ed.), *Hegel's Philosophy of Right* (pp. 43-72). New York: Wiley-Blackwell.

Guyer, Paul. (2006) *Kant*. New York: Routledge.

Hernandez, Jill. (2010) "Impermissibility and Kantian Moral Worth," *Ethical Theory and Moral Practice* Vol. 13, No. 4: 403-419.

Kahn, Samuel. (2019) "Defending the Traditional Interpretations of Kant's Formula of a Law of Nature." *Theoria* Vol. 66, Issue 158: 76-102.

O'Neill, Onora. (2013) *Acting on Principle*. Cambridge: Cambridge University Press.

ⁱ O'Neill's claim about extrapolation is interesting in part because it seems to suggest that the first two steps of her proposal are *not* extrapolations: that, according to O'Neill, Kant thought that the deontic status of a maxim should be determined by testing it along with its contrary. One might wonder whether the extrapolation lies mainly in recognizing that category 3 is nonempty. It is unclear, and I am not going to engage in any of the textual questions that this raises.

ⁱⁱ There is a great deal of dispute regarding the correct interpretation of the universalizability tests and I do not want to take a stand on that here. For a (critical) survey of five recent interpretations of the universalizability tests, see my (Kahn, 2019).

ⁱⁱⁱ Note that the scope of the modal term in this principle is the conjunction A *and* B. Note also that, technically, a weaker principle would suffice for my purposes (namely: if it is obligatory to A and if it is impossible to A and to B, then it is not obligatory to B). Both of these principles (the one in the main text and the weaker one in this note) put pressure on those who attempt to solve the problem of moral dilemmas by rejecting the principle of agglomeration. So perhaps some, on reflection, would reject this principle.

^{iv} Some might argue that this thesis is analytic, that taking the necessary means to an action is *constitutive* of performing that action and, thus, that the obligatoriness of the necessary means to an action is not *derivative* of the obligatoriness of the action but rather *constitutive* of it. I suspect that epistemic considerations might pose a problem to this idea, but I cannot pursue this line of thought there.

^v An alternative proposal would be to redefine the concept of a maxim contrary so that a maxim has only one contrary. For example, one might assert that two maxims are contraries if but only if (a) their UTCs are contraries *and* (b) they have the form "to A" and "not to A."

I think the easiest way to see the problem with this proposal is to focus on O'Neill's non-reciprocal action maxims. As noted above (in section II), non-reciprocal action maxims, like maxims E and F, are, according to O'Neill, supposed to be contraries. But on this new definition, non-reciprocal action maxims no longer would be so because they do not have the form "to A" and "not to A"; the contrary of maxim E on this new definition would be "to sell toy trains *or* never to buy them" (rather than "to sell toy trains *and* never to buy them"). The difference is subtle but important. The reason it is important is that the disjunctive maxim would be universalizable on O'Neill's account, whence it would follow (because maxim E is not universalizable) that the disjunctive maxim is obligatory whereas maxim E is impermissible. Moreover, it is easy, once one sees this, to derive many other false positives and false negatives using this new definition of a maxim contrary. The underlying problem may be formulated as a dilemma: on one horn, there are many maxim pairs of the form "to A" and "not to A" that agents never deliberate about (and there is no necessity, neither logical nor metaphysical nor physical nor psychological, that an agent will adopt either one); while on the other horn, agents' deliberative options frequently are not contraries in any plausible sense, neither O'Neill's nor this other version. I would like to thank Richard Galvin for pressing me to clarify my thoughts about this.