Positive duties, maxim realism and the deliberative field, forthcoming in *Philosophical Inquiry*

I.0 Introduction

My goal in this paper is to show that it is not the case that positive duties can be derived from Kant's so-called universalizability tests. I begin by explaining in detail what I mean by this and distinguishing it from a few things that I am not doing in this paper. After that, I confront the idea of a maxim contradictory, a concept that is advanced by many commentators in the attempt to derive positive duties from the universalizability tests. I explain what a maxim contradictory is and how the concept is used to derive positive duties. Then I argue that the notion of a maxim contradictory presupposes an objectionable form of maxim realism. I move from there to the idea of a maxim contrary and the deliberative field. These two ideas are used in tandem by commentators who do not appeal to maxim contradictories. I explain how these concepts are used to derive positive duties and then I argue that there is a systematic error in the derivations that enables one to see that they cannot work.

I.1 More detailed explanation of thesis

An agent has a positive duty if but only if there is some action or maxim (i.e., principle of action) that it is obligatory for him/her to perform or adopt. This definition can be given equivalently using impermissibility: an agent has a positive duty if but only if there is some action or maxim that it is impermissible for him/her to omit or not to adopt. If there is a particular action that it is impermissible for an agent to omit, then

1

there is an action that, if performed, one could point at and say, "if s/he had not done *that*, s/he would have been behaving impermissibly." If there is a particular maxim that it is impermissible for an agent not to adopt, then there is a maxim one can point at and say, "if s/he has not adopted *that* maxim, s/he is behaving impermissibly." For example, if it is obligatory for agents to adopt the maxim "sometimes to help others," then agents have a positive duty to adopt this maxim.

The concept of a positive duty can be made clearer by distinguishing it from the concept of a negative duty: an agent has a negative duty if but only if there is some action or maxim that it is obligatory for him/her to omit or not to adopt. This definition, too, can be given equivalently using impermissibility: an agent has a negative duty if but only if there is some action or maxim that it is impermissible for him/her to perform or adopt. For example, if it is impermissible for agents to adopt the maxim "to enslave all humans whose skin color differs from mine," then agents have a negative duty not to adopt this maxim. To sum up, a positive duty is a duty to adopt a maxim; a negative duty is a duty not to adopt a maxim.[i]

When I say that it is not the case that positive duties can be derived from *Kant's so-called universalizability tests*, I am referring to two formulations of the Categorical Imperative (CI): the formula of universal law (FUL) and the formula of a law of nature (FLN).[ii] FUL is the first formulation of the Categorical Imperative introduced by Kant in the *Groundwork for the Metaphysics of Morals*.[iii] It runs as follows:

> Act only according to that maxim by which you can at the same time will that it should become a universal law. (4:421)

The test proposed here is to see whether I can will that my maxim (in any given instance) can become a universal law. That is, according to FUL, willing the universalized version of my maxim ought to be consistent (in some as yet unspecified way) with willing the maxim itself.[iv]

In part II of the *Groundwork for the Metaphysics of Morals*, Kant introduces FLN. It runs as follows:

> Act as if the maxim of your action should through your will become a universal law of nature. (4:421)

The main difference between FUL and FLN is in willing something to be a universal law as opposed to willing it to be a universal law of nature. A universal law (*simpliciter*) is a principle that perfectly rational agents would follow without exception (i.e., one that imperfectly rational agents ought to follow). A law of nature is a causal principle that agents follow by necessity without exception insofar as they are members of a system of nature. Agents are to regulate their conduct "as if" they could raise the principles of their wills up to principles of the system of nature to which they belong. Kant seems to regard these two formulations as equivalent (for finite, embodied creatures like humans) in the sense that they will agree about the permissibility of a given maxim,[v] and for the purposes of this paper there is no need to differentiate between them.

When I say that it is not the case that positive duties *can be derived* from Kant's so-called universalizability tests, I am making a very specific claim. As will be argued below, the universalizability tests give marks that any maxim must have if that maxim is permissible. The marks have to do with whether an agent can will the maxim as a univer-

sal law or as a law of nature.[vi] Thus, the universalizability tests give informal decision procedures that can be used to determine whether a given maxim is permissible. Of course, the application of these decision procedures requires judgment. But the point is that the universalizability tests give procedures for determining negative duties: there is no analogous procedure for determining whether a given maxim is obligatory (a positive duty).

I.2 Some things I am not doing

In the previous subsection, I tried to explain my use of three pieces of jargon: (1) the positive/negative distinction, (2) the universalizability tests and (3) the sense in which I argue that positive duties cannot be derived from the universalizability tests. In this sub-section, I shall try to sharpen these three explanations. I shall do so by noting some potential misunderstandings that could arise for each one about the thesis of this paper. I have culled the discussion in this subsection from objections I have received. Feel free to skip this subsection if the previous one was sufficiently clear.

The positive/negative duty distinction is sometimes put together with other distinctions discussed in Kant studies. For example, Allison argues that in Kant's ethics "in addition to the narrow, perfect duties to respect ends in the negative sense there are also broad, imperfect duties to adopt ends in the positive sense."[vii] In this passage, Allison appears to be treating three distinctions as equivalent: the positive/negative duty distinction, the perfect/imperfect duty distinction and the narrow/broad duty distinction.

Kant uses both the perfect/imperfect distinction and the narrow/broad distinction in the *Groundwork for the Metaphysics of Morals*.[viii] For example, in part II of the text, Kant glosses "strict or narrower" duty as "unremitting" duty, to be contrasted with "meritorious" duty,[ix] and he argues that this distinction tracks a distinction in the way in which a maxim can fail the universalizability tests (4:424). He also appeals to the distinction between perfect and imperfect duties, explaining how he understands the concept of a perfect duty in a footnote: "I understand here by a perfect duty one that admits no exception in favor of inclination" (4:422n). Kant's use of these and other distinctions to carve up the landscape of duties can be confusing, and many scholars have written about how they ought to be understood.[x] Taking them to be equivalent therefore threatens to distract from the main argument of the paper, which does not necessitate delving into these debates.

The universalizability tests are sometimes put together or contrasted with other formulations of the CI in discussions of positive duties. For example, Allison seems to think that positive duties cannot be derived from any formulation of the CI[xi] without the assumption of transcendental freedom.[xii] Moreover, there are complicated questions about whether any or all of the formulations of the CI are equivalent.[xiii] I shall not have anything to say about these questions in this paper. My discussion is confined to the universalizability tests, and if my argument works, then even assuming transcendental freedom will not suffice for the derivation of positive duties from the universalizability tests.

In addition to these issues about which formulations of the CI are equivalent, there are questions about how to interpret the universalizability tests.[xiv] I shall not be tak-

ing a stand on these questions. I also shall not be taking a stand on whether FUL and FLN are equivalent.[xv] My discussion of the universalizability tests is based on the idea that they are tests of permissibility. To see why I think this, note that FUL starts with the expression 'act only' ('*handle nur*') and FLN starts with the expression 'act as if' ('*handle so*'). If I tell you to act only/as if according to maxims that have some property X, then I am telling you that all of your maxims better have that property. In other words, the universalizability tests say that if an agent acts on a maxim that can*not* be willed at the same time as a universal law (of nature), then that agent is acting impermissibly, and if an agent acts on a maxim that *can* be willed at the same time as a universal law (of nature), then that agent is acting permissibly.[xvi] If this is correct, then to show that a maxim is (not) universalizable does not show that there is a positive duty to adopt that maxim. Thus, some additional argument is needed to derive positive duties from the universalizability tests. My goal is to show that no such argument works.[xvii]

My claims about derivations might be misunderstood. I am not going to argue that (1) in order to determine whether a given maxim is obligatory one would have to go through an infinite number of other maxims and show that each one is impermissible, but (2) there is no fixed enumeration of maxims, so (3) determining whether a given maxim is obligatory using the universalizability tests is impossible. Similarly, I am not going to argue that (1) even if there were such a fixed enumeration of maxims, one never shall be in a position to test an infinite number of maxims, so (2) determining whether a given maxim is obligatory using the universalizability tests is impossible. Rather, I am going to argue that even if there were such a fixed enumeration and even if one were in a position

to test an infinite number of maxims (there is no such fixed enumeration and one never shall be in a position to test an infinite number of maxims, but even if, *per impossible*, these conditions were to obtain) the best one could do is narrow things down to a set of maxims (plural) and say that it is obligatory that an agent adopt a maxim from this set. But that is not a positive duty.

Finally, I should note that this is not an exegetical thesis. I am not trying to show that the standard account of how to derive positive duties from the universalizability tests in Kantian ethics is not textually warranted. I also am not trying to show that positive duties are in some way superfluous. My project here is to show simply that positive duties cannot be derived from the universalizability tests.

II.0 Maxim realism

Perhaps the most popular way to argue for positive duties using the universalizability tests is to appeal to the notion of a maxim or an action contradictory. I shall begin by reproducing some of these arguments. Then I shall explain the basic idea behind these arguments in my own words before going on to explain why they do not work.

II.1 Commentators appeal to maxim contradictories

In a discussion of the derivation of the duty of beneficence, Guyer argues as follows:

> …just as the falsity of one statement implies the truth of its contradictory, e.g. 'It is false that it is raining today' implies 'It is true that it is not raining today,' so it would seem that the impermissibility of adopting one maxim does imply the necessity of adopting its contradictory, thus the impermissibility of the maxim of never helping anyone does imply the necessity of the maxim of helping someone sometime.[xviii]

Herman expresses exactly the same line of thought:

> …if 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.'[xix]

Korsgaard uses 'opposite' instead of 'contradictory,' but otherwise the argument is identical:

> …if a maxim fails the [universalizability] test…we get a duty—the duty of doing the opposite of what the failed maxim says, or, more precisely, of adopting the opposite of what the maxim says as a law.[xx]

Korsgaard then goes on to illustrate this argument with a maxim about false promising.[xxi]

This is relevant because of what was said above about the positive/negative distinction and the perfect/imperfect distinction: in Kant studies, the duty of beneficence (mentioned in the quotations from Guyer and from Herman, respectively) is traditionally taken to be an imperfect duty whereas the duty to keep one's promises is traditionally taken to be a perfect duty. Thus it may be seen that the positive/negative distinction cuts across the perfect/imperfect distinction in these discussions.[xxii]

Harrison expresses the idea in a discussion of how to interpret the results of submitting a maxim to the universalizability tests:

> 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 universifiable…(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.[xxiii]

Stratton-Lake, like Harrison (in the last sentence of this quotation), focuses on actions, but his thought seems to be more or less the same as that of the other commentators in the passages expressed above:

> …in relation to obligatory acts, we can understand it as doing what one should just because the maxim of the contradictory act cannot be universalized.[xxiv]

Now I shall summarize the arguments being made by these commentators in my own words.

II.2 The derivation of positive duties using maxim contradictories

In general, two predicates X and Y are contradictories if but only if for all Z, Z is X or Z is Y but not both. For example, alive and lifeless are contradictories: everything is either alive or lifeless but nothing is both alive and lifeless (at the same time, anyway). Applied to maxims, the scope of contradictories is narrowed to the universe of agents. X and Y are maxim contradictories if but only if X and Y are maxims and for all A, if A is an agent, either A has X or A has Y but not both. The notion of a maxim contradictory gets its intuitive pull from the following line of thought. For any given action description and for any given agent, that agent either would perform that kind of action or s/he would

not. If the former, then s/he has the maxim sometimes to perform that kind of action; if the latter, then s/he has the maxim never to perform that kind of action. Hopefully an example will make the idea clear. Consider the maxim, "to tell a lying promise." Either you would tell a lying promise or you would not. If the former, you have the maxim "sometimes to tell a lying promise"; if the latter, you have the maxim "never to tell a lying promise." Other ways to cash out the notion of maxim contradictories include appeal to values (if an agent never would X, then s/he has the maxim "unconditionally not to X") and using phrases like "on no account." These are all after the same idea.[xxv]

Given the concept of a maxim contradictory, the universalizability tests can be used to derive positive duties as follows. Run the universalizability tests on the two maxims in a maxim-contradictory pair. For ease of reference, I shall call the original maxim O_m and I shall call its contradictory C_m. If O_m is universalizable and C_m is not, then O_m is permissible and C_m is not. Because (by definition of a maxim contradictory) all agents adopt either O_m or C_m, it follows immediately that O_m is obligatory and, therefore, agents have a positive duty to adopt O_m (and a negative duty not to adopt C_m). If O_m is not universalizable but C_m is, then the same line of reasoning will show that agents have a positive duty to adopt C_m (and a negative duty not to adopt O_m).[xxvi]

This can be illustrated using the maxim-contradictory pair appealed to in the first two quotations above (Guyer and Herman). The maxim "sometimes to help others" passes the universalization tests and therefore is permissible, whereas the maxim "never to help others" does not pass the universalization tests and therefore is impermissible.[xxvii] But all agents adopt either the maxim "sometimes to help others" or the maxim "never to

help others," for these are maxim contradictories. So all agents have a positive duty to adopt the maxim "sometimes to help others" (and a negative duty not to adopt the maxim "never to help others"). A similar line of thought will show that agents have a positive duty to adopt the maxim "never to tell a lying promise" (and a negative duty not to adopt the maxim "sometimes to tell a lying promise").[xxviii]

As hinted at in the quotation from Harrison, the procedure for forming a maxim and its contradictory is relatively straightforward. On the one hand, append the word 'sometimes' to a given action description; on the other hand, append the word 'never' to the action description. For example, if the action description is "to help some others," then the maxim-contradictory pair becomes "sometimes to help some others" and "never to help some others" (or, more colloquially, "never to help anyone"). If, alternatively, the action description is "to tell a lying promise," then the maxim-contradictory pair becomes "sometimes to tell a lying promise" and "never to tell a lying promise."

Now I want to emphasize one point again before I explain why I take this account not to work. If maxims X and Y are contradictories, then for all agents A, A has X or A has Y but not both (at the same time). This is why Guyer, for example, asserts that the impermissibility of a maxim logically entails the obligatoriness of adopting its contradictory in just the same way that the falsity of one statement logically entails the truth of its contradictory. But perhaps at this point you (the reader) already can guess where I am going when I say that this presupposes an objectionable form of maxim realism, and perhaps you will say that the account requires only a small tweak to avoid this presupposition: instead of committing to the idea that all agents have either the maxim "sometimes

11

to X" or the maxim "never to X," it will suffice to say that *when an agent is deliberating about whether to X*, s/he must adopt one of these maxims.

This, I think, is a good objection. In fact, this is the basic insight that underlies the account I argue against in the next section of this paper. The crucial point for this section is that this tweak involves changing the argument that is being made by the commentators cited above. In particular, it involves jettisoning maxim contradictories in favor of maxim contraries. The next section will begin by citing commentators who do this and then ex-plaining the argument in my own words before explaining why it does not work. But for now, I am going to focus on the account rehearsed above and, in particular, on the idea that the notion of a maxim contradictory presupposes an objectionable form of maxim realism.

II.3 Why I think the notion of a maxim contradictory presupposes an objectionable form of maxim realism

The reason I say that the notion of a maxim contradictory presupposes an objec-tionable form of maxim realism is that it presupposes that maxim-contradictory pairs are out there, as if on a list, and all agents (necessarily) have one or the other from every pair of maxims on this list associated with them (somewhat like Leibnizean complete individ-ual concepts). I shall begin by focusing on the acquisition process. Becoming an agent means different things in different contexts. In Kantian ethics, becoming an agent means acquiring certain capacities.[xxix] Principal among those is the capacity for self-governance,

the capacity not to be pulled around by one's impulses: the capacity not to be a wanton. The predicate "is an agent" is, in Kantian ethics, vague. It is interesting in that regard: people do not begin life as agents, and many people do not end life as agents, either.[xxx] Somewhere around age 18-21,[xxxi] a child matures into an agent; some time later, that agent might have a "second childhood." But the point is that if for all maxim contradictory pairs, all agents have one of the two, then upon attaining agency, a child suddenly acquires an infinite set of maxims. In particular, for every action description A (of which there are, presumably, infinitely many), an agent acquires either the maxim "sometimes to A" or the maxim "never to A."

Now it might be objected that I am being unfair. Becoming an agent is a slow process. My remarks about the predicate "is an agent" bear witness to this to some degree. So there does not need to be anything sudden about the acquisition of this infinite set of maxims. Moreover, there is nothing counterintuitive about saying that an agent has an infinite set of maxims. Agents are able to react to an infinite number of stimuli; agents can form an infinite number of sentences; agents have the concept of infinite; why not say that agents have an infinite set of maxims?

But both of these replies miss the mark. The deeper problem with saying that upon maturity an agent acquires an infinite set of maxims has nothing to do with the duration of the acquisition process or with the number of maxims in the set. The problem is that in addition to acquiring maxims about activities in which s/he has engaged—maxims about star-gazing, about typing, about writing, about reading—the agent acquires maxims about things s/he never in a million years would contemplate doing. S/he acquires max-

13

ims about cannibalism, about cow tipping, about botulism. S/he acquires maxims about things it is not appropriate to mention in an academic paper, and s/he acquires maxims about things it is not appropriate to mention out of an academic paper. Moreover, it must be remembered that maxims are taken to be subjectively normative principles of action: they are rules by which an agent governs him/herself.[xxxii] So to say that agents acquire maxims about activities they have not contemplated is somewhat odd. The fact that there is no plausible explanation of the acquisition process for such maxims only highlights the problem.

But perhaps what I just said about the nature of maxims exposes a problem with my attack on maxim contradictories: it might be objected here that there is a distinction between maxims, as principles of action, and rules.[xxxiii] Maxims are *general* principles of action; rules are *specific* intentions. For example, "sometimes to help others" certainly is a maxim (at least according to those against whom I am arguing). But "under no condition to help Hilly Flitcraft if he asks me to lend him $4.23 on the third Monday of March 2001" probably is "merely" a rule, a specific intention.[xxxiv] The idea behind raising this objection here is that many of the putative maxims that I am supposing a person will acquire upon reaching agency are not really maxims. So the supposed counterintuitive implications of the maxim realism presupposed by the notion of a maxim contradictory (*viz.*, that agents acquire, as if by magic, maxims about things they never in a million years would contemplate doing) do not go through.

But this objection, like the last, misses the mark. The trouble with this objection is not that the distinction between maxims and rules is hazy at best.[xxxv] That certainly seems

to be true, but that is not why this objection misses the mark. This objection misses the mark because the maxims I am talking about are all quite general. They are maxims just as general as the maxim "sometimes to help others." For example, "sometimes to eat others," "sometimes to filet others," "sometimes to gut others," "sometimes to disembowel others," "sometimes to eviscerate others," "sometimes to fry others," "sometimes to bake others"—and the list could go on. Of course, there are agents who have had some of these (perhaps most of these) maxims. But I have listed only a few of the tamer ones. And the point is that most people have not adopted any of these maxims, so it is odd to think that they acquire them simply by virtue of reaching agency.

It should be noted (by way of clarification) that the idea here is not that agents fail to fulfill their duties in some counterintuitive way if the standard account of positive duties is accepted. That is, the idea is not that the standard account yields false positives (or false negatives) for positive duties. The standard account *might* yield false positives (or false negatives) for positive duties, but that is not the point I am trying to make. Rather, the point I am trying to make is that the standard account relies on the notion of a maxim contradictory, but the notion of a maxim contradictory presupposes a form of maxim realism that is objectionable insofar as it entails that agents have all sorts of maxims that they do not have. The point I am making is independent of "becoming an agent." I am pointing out quite simply that the standard account presupposes that there is a list, so to speak, of maxims and their corresponding contradictories, that every agent (by definition of a maxim contradictory) has one maxim from every pair on the list—and that this entails that agents have maxims that they do not have. Again, consideration of the fact that

agents gain these maxims (by *fiat*) upon reaching agency only serves to make the point more vivid.

It might be objected that agents really do have all of these maxims and that this can be seen by engaging in counterfactual reasoning.[xxxvi] The idea underlying this objection can be seen by example. Consider the maxim, "sometimes to bake others." To see whether an agent has adopted this maxim, one asks whether there are any circumstances in which the agent in question would bake someone. Of course, it would not count for current purposes if there are circumstances in which the agent in question would bake someone inadvertently. For example, if unbeknownst to the agent someone climbed into the oven while s/he was out and then s/he returned to make some cookies and turned the oven on to preheat—this would not be relevant for current purposes. This would not show that the agent has (or does not have) the maxim "sometimes to bake others." But if there are circumstances in which the agent intentionally would bake someone, then s/he has the maxim, "sometimes to bake others."

Note the claim here. The claim is not that the results from these thought experiments demonstrate whether an agent *would adopt* a given maxim. The claim is that the results from these thought experiments demonstrate whether an agent *already has adopted* a given maxim. Thus, a proponent of the standard account might ask, "what more is there to figuring out an agent's maxims than seeing how an agent would behave in counterfactual circumstances?"

But there *is* something more, and one need not articulate a complete account of what it is to adopt a maxim to see why this question is problematic. I take it as uncontro-

16

versial that in order for an agent to have adopted a maxim, it is necessary that s/he be able to articulate that maxim given his/her knowledge of the world.[xxxvii] This is not to say that an agent actually must articulate a maxim to adopt it: I am not claiming, for instance, that to adopt a maxim, an agent must engage in conscious deliberation about that maxim. Similarly, this is not to say that an agent actually must have the maxim present before his/her mind's eye: I am not claiming, for instance, that to adopt a maxim, an agent must be thinking about that maxim (and/or be aware that s/he is thinking about that maxim).

The claim is quite minimal: if an agent is unable to articulate a maxim given his/her knowledge of the world (so if an agent lacks certain concepts in the maxim) then s/he cannot have adopted it. S/he might adopt it if s/he gains additional knowledge of the world. And, indeed, one might make plausible conjectures about *whether* s/he would adopt it *if* s/he gained that additional knowledge of the world. But such conjectures would demonstrate something about his/her character, not something about whether s/he already has adopted the maxim in question. It would not demonstrate whether s/he already has adopted the maxim in question because it is already known that s/he has not, and this is already known because s/he is incapable even of articulating the maxim or understanding it if it is articulated to him/her.

As an example, consider the maxim "sometimes to kick bulldogs" and its contradictory, "never to kick bulldogs." The point I am making is that there are agents who have adopted neither of these maxims. And this can be seen by appeal to the fact that there are agents who do not know what a bulldog is. Indeed, there are agents who grew up in times before bulldogs had evolved. And once this is accepted, it can be seen why

the standard account does not work: its attempt to derive positive duties hinges on a notion (*viz.*, that of a maxim contradictory) that presupposes an objectionable form of maxim realism.[xxxviii]

III.0 Maxim contraries and the deliberative field

As noted above, proponents of the argument explored in the previous section could jettison the notion of maxim contradictories in favor of maxim contraries and, it seems, still derive positive duties from the universalizability tests. I shall begin, as before, by reproducing some of these arguments. Then I shall explain them in my own words before going on to explain why I think they do not work.

III.1 Commentators appeal to maxim contraries and the deliberative field

Some commentators appeal explicitly to the notion of maxim contraries:

> …questions have been raised about whether FUL and/or FLN really do yield all our duties…several [commentators] have argued that the universalizability test[s] gives rise only to negative and not positive duties… [this] objection seems incorrect: 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.[xxxix]

Other commentators make roughly the same point but without an explicit appeal to maxim contraries:

> Kant does derive obligatory ends from the Formula of Universal Law, but he does it by a curiously roundabout procedure in which someone is imag-

18

ined 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 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…[xl]

Similarly, in a discussion of an agent who contemplates adopting a maxim of indifference, Engstrom argues as follows:

If this way of thinking were universal, others would judge their happiness but not his to be simply good, and if he were to agree with them in this judgment, he would likewise regard their happiness but not his to be simply good. He would thus deem good his happiness but not theirs and also their happiness but not his. So his maxim contradicts its form. And since all persons necessarily wish for their own happiness, his only way of avoiding this inconsistency is by extending his judgment of the goodness of his own happiness so that it includes that of others…[xli]

Engstrom enlarges on this argument a few pages later:

This man does not ask "What concern is it [*viz.*, the happiness of others] of mine?" simply out of the blue. The question, along with its hint of irritation, arises in response to a thought that is already present, if only obscurely, in his own mind—namely that others' happiness *is and ought to be* a concern of his. Seeing others who have to struggle with great hardships arouses his own obscure practical understanding that their happiness is a matter of concern, and not just to them, and this stimulation leads to the further reflections described in the example.[xlii]

I now shall try to summarize these arguments in my own words.

III.2 The derivation of positive duties using maxim contraries and the deliberative field

In general, two predicates X and Y are contraries if but only if for all Z, it is not the case that Z is X and Z is Y. For example, blue all over and green all over are con-

traries: nothing is both blue all over and green all over (at the same time). But some things are neither blue all over nor green all over. Predicates that are contradictories are also contraries. But the reverse does not hold.

Applied to maxims, the scope of contraries (like the scope of contradictories) is narrowed to the universe of agents. X and Y are maxim contraries if but only if X and Y are maxims and for all A, if A is an agent, it is not the case that A has X and A has Y. As with the definition of maxim contraries, some tightening here is needed in order to account for the possibility of changing maxims through time. Moreover, although it is less important for current purposes, it is worth noting that if by chance no actual agent ever has had both the maxim "sometimes to help others" and the maxim "sometimes to eviscerate others," it would not be said that these two maxims are contraries. One easy way to get around this problem would be to say that the universe of agents referred to in the definition of a maxim contrary quantifies over all possible agents. However, I shall not pursue this line of thought.

Now the argument made for positive duties using the concept of a maxim contradictory cannot be applied using the concept of a maxim contrary. To see why, it will suffice to note that if X and Y are contraries, then an agent might not adopt either X or Y, so if X is permissible and Y is impermissible, it does not follow that agents have a positive duty to adopt X (and a negative duty not to adopt Y). This is the point behind the claim in the second excerpt about the fact that this argument "does not show that there would be a moral failing if the agent merely unthinkingly neglected rather than rejected these ends": because it is not the case that all agents adopt one or the other of these maxims, if the

agent never deliberates about them to begin with (unthinkingly neglects them), there will be no moral failing. Moreover, there is another point that requires consideration: maxim contradictories come in pairs; maxim contraries do not. For example, the following three maxims might be considered contraries: "sometimes to help others but sometimes not to help others," "never to help others" and "always to help others." In order to illustrate the idea behind the argument I want to attack, it will be helpful, initially at least, to confine consideration to maxims constructed using the method described in the previous section for maxim contradictories: take a general action description and append to it 'sometimes' on the one hand and 'never' on the other.

Moving ahead with the maxim contrary construction procedure just described, the crux of this new argument is the idea that *when an agent is deliberating about whether to X,* s/he must adopt the maxim "sometimes to X" or the maxim "never to X." It is here that the concept of the deliberative field enters the fray.[xliii] The concept of the deliberative field does not allow of a nice definition in the way in which the concept of a maxim contrary (and a maxim contradictory) does.[xliv] But I shall try to make the concept clear by means of some general remarks and some examples.

The concept of the deliberative field is supposed to capture what an agent is aware of when s/he is engaging in deliberation. Presumably an agent need not be aware that s/he is aware of something for that thing to be within his/her deliberative field. Similarly, an agent need not be deliberating consciously about something for that thing to factor into his/her deliberative field. But if I am deliberating about whether to catch a train or a bus and someone points out to me that I should take the train because there will be a lot of

traffic on the roads today, then that person is raising a concern about traffic on the roads to my deliberative field. Whether that person succeeds depends on whether I am interested in making it to my destination within a certain amount of time.

So how does the deliberative field help here? The basic idea is something like this. Insofar as an agent is considering whether to engage in some kind of action A, s/he has Aing within his/her deliberative field. But if s/he is considering whether to engage in some kind of action A, s/he must adopt a maxim contrary about A. To refer back to the second excerpt above, it will be impossible for the agent "unthinkingly" to neglect A in such a situation. This is impossible precisely because the agent is thinking about A, and so the agent must adopt either the maxim "sometimes to A" or the maxim "never to A." It follows immediately that if the maxim "never to A" is impermissible, then there *would* be a moral failing (in just the way described in the second excerpt above) if s/he fails to adopt the maxim "sometimes to A." Thus, the agent has a positive duty to adopt the maxim "sometimes to A." The difference between this argument and the argument made using maxim contradictories is that this argument hinges on a hypothetical: *if the agent is deliberating about A.* But perhaps an example will make this argument clearer. (Again: I am still simply trying to illustrate the argument I want to attack; I have not gotten to my attack yet.)

Suppose I am considering whether to break a promise. Then the two maxim contraries relevant to this case are "sometimes to break promises" and "never to break promises." Insofar as I am engaged in deliberation about whether to break a promise, I must adopt one of these two maxims. But the maxim "sometimes to break promises" is

impermissible. Therefore, I ought to adopt the maxim "never to break promises": I have a positive duty to adopt this maxim. Alternatively, suppose I am considering a maxim of indifference: I am considering whether to help others. Then the two maxim contraries relevant to this case are "sometimes to help others" and "never to help others." Insofar as I am engaged in deliberation about helping others, I must adopt one of these two maxims. But the maxim "never to help others" is impermissible. Therefore, I ought to adopt the maxim "sometimes to help others": again I have a positive duty (this seems to be Engstrom's line of thought when he says of the man in his example, "his only way of avoiding this inconsistency [namely, that associated with a maxim of indifference] is by extending his judgment of the goodness of his own happiness so that it includes that of others").xlv

The last step in this argument is to discharge the hypothetical. The positive duties derived by this method are to some extent situation-specific: they depend on what an agent is deliberating about doing. But it is difficult to imagine an agent who never has considered or will consider whether to break a promise, and it is perhaps even more difficult to imagine an agent who never is confronted with a situation in which s/he has the opportunity to help others. This is the point behind the claim about "the pervasiveness of these ends in the moral life" in the second excerpt above. It also seems to be the point behind Engstrom's claim that the man in his example does not question whether a maxim of indifference is permissible "out of the blue." Engstrom defends Kant's claim that happiness is an end for every human being, and so it is only natural that an agent should deliberate about whether s/he ought to help others in need.xlvi Thus it may be seen that by

23

making its positive duties conditional upon what an agent deliberates about, this account avoids the problems I was raising for maxim contradictories in the previous section. Moreover, in many cases (promise keeping and helping others, for example) the kinds of deliberation are so pervasive that most agents will have to adopt these maxims early on in life.

III.3 Why I think the appeal to maxim contraries and the deliberative field does not work

Now I shall explain why I think this argument does not work. Consider again the claim that if an agent is deliberating about whether to A, then s/he must adopt the maxim "sometimes to A" or the maxim "never to A." I contend that this claim is false. The claim is not false merely because, for example, an agent might be interrupted in his/her deliberations before s/he has a chance to adopt either of these maxims. That kind of issue is (presumably) soluble. The argument might need some retooling. But that is not what I want to focus on. I contend that the claim is false for a deeper reason: the claim is false because an agent could adopt a completely different maxim about Aing. In point of fact, there is an untold number of different maxims about Aing that an agent could adopt. Let me explain.

Suppose again that I am considering whether to break a promise. It certainly does seem possible for me to adopt the maxim "sometimes to break my promises." Similarly, it seems possible for me to adopt the maxim "never to break my promises." And these two

24

maxims do seem to be mutually exclusive. The problem that I am raising is that they are

not exhaustive. I also could adopt the maxim "not to break my promise *now*." And impor-

tantly, I could adopt this maxim without any consideration about what I might do about

promises in the future.

Of course, if my objection went no further than this, then I doubt I would win

much support. No matter where one stands on the maxim/rule debate (discussed, briefly,

in the previous section), "not to break my promise *now*" is likely to look too much like a

specific intention to count as a maxim. But that is easily remedied: it is a consequence of

the fact that the example is under-described. Suppose that in considering whether to break

a promise, I am considering whether to break a promise to meet a long-standing chess

rival of mine for a weekday lunch and game of blindfolded chess. Even if it is assumed

that the maxim "to break promises to meet up with people" in some sense implies the

maxim "sometimes to break my promises" so that in adopting the former, an agent *eo

ipso* adopts the latter[xlvii] there is a whole manifold of maxims that could be adopted here

that would imply neither the maxim "sometimes to break my promises" nor the maxim

"never to break my promises"—to wit, "not to break promises to meet up with people,"

"not to break promises to play chess," "not to break promises to do things on weekdays,"

etc.

Note that I am not making a claim about false positives or false negatives. I am

not claiming, for example, that some of these maxims pass the universalizability tests

even though they are impermissible (perhaps because they are so specific).[xlviii] I am

claiming that these maxims are permissible and simply assuming (for the sake of argu-

ment) that the universalizability tests would show this. After all, if "never to break my promises" is permissible, presumably "not to break promises to meet up with people" is permissible. But if these maxims are permissible, then it follows that there is no positive duty to adopt the maxim "never to break my promises," not even from within the deliberative field.

However, someone might object to this argument on the following grounds: someone might argue that this is an artifact of my using "promise breaking" rather than "promise keeping." If not breaking a promise is the same thing as keeping a promise, then each of the maxims I listed at the end of the paragraph above could be written in terms of promise keeping. Thus, rather than providing a good objection to the argument I am confronting, it might be argued that I have played a sort of parlor trick—some clever word play and nothing more.

But this objection can be laid to rest quite quickly. Suppose the positive duty is taken to be "always to keep my promises" rather than "never to break my promises." The relevant contrary is, then, "sometimes not to keep my promises." These two maxims again seem mutually exclusive. But, again, they are not exhaustive, and the reasoning does not change: I also could adopt the maxim "to keep my promise *now*." Naturally I do not expect many people to accept that *this* is a maxim rather than a specific intention. But as I already have explained, this is easily remedied: it is a problem with the fact that the example is under-described. A whole manifold of maxims, analogous to those gestured toward above, would be easy enough to cook up. The point is that if I decide (not) to keep or break a promise on one occasion, this does not imply that I have a fully worked out

and principled stance on when I shall keep or break promises and when not. I simply might have come to the conclusion that although it *might* be permissible to keep or break a promise under some circumstances (it equally well might not—it is hard to say, and I do not need to judge about that now: I shall cross that bridge if and when I come to it) it is certainly not permissible to do so under these circumstances, so I am not going to do it.

The problem that the appeal to maxim contraries faces is that there is always more than one permissible maxim that an agent can adopt in any given instance about any given action: never will it be possible using a permissibility test like the universalizability tests to arrive at a positive duty because there is always a plurality of permissible maxims from which to choose. Again: if the maxim "never to break my promises" is permissible, then so (*a fortiori*) is the maxim "never to break a promise I made in secret," and it follows that neither of these two maxims is obligatory. Moreover, it is quite easy to imagine an agent who thinks that the maxim "never to break a promise I made in secret" is permissible but who remains agnostic about the maxim "never to break my promises." So it is quite easy to imagine an agent who actually would adopt one without adopting the other. And lest it be thought that this works only in one direction, I hasten to point out that it is just as easy to imagine an agent who adopts the maxim "never to break my promises" without considering promises made in secret, an agent who *reconsiders* his/her commitment to the maxim "never to break my promises" when s/he begins to contemplate breaking a promise made in secret. One might call such an agent Actutilitarius, and one might argue that Actutilitarius has adopted the maxim "never to break my promises" without adopting the maxim "never to break a promise I made in secret."

27

However, it might be objected that in focusing on the putative positive duty about promises, I have missed the mark. At least some of the commentators reproduced above might concede that there is no positive duty to adopt a maxim "never to break my promises." They argue for the positive duty to adopt a maxim "sometimes to help others." Moreover, at least one of the arguments I made above does not obviously apply to this maxim. Above, I claimed that if the maxim "never to break my promises" is permissible, then so (*a fortiori*) is the maxim "never to break a promise I made in secret." What is the parallel for the maxim "sometimes to help others"? It is unclear. And so even if my objection to the derivation of a positive duty regarding promises goes through, more work must be done to show that the derivation of the positive duty of beneficence (and others like it) does (do) not work.

There are two lines of attack that I would like to press against the derivation of the duty of beneficence. First, as with the duty never to break promises, the maxim "sometimes to help others" and the maxim "never to help others," although certainly mutually exclusive, are not exhaustive. On the one side, consider the maxim "sometimes to help others." If this maxim is permissible, then there (perforce) is some (even if only slightly) more specific maxim about helping others that is permissible. On the other side, consider the maxim "never to help others." Although this maxim is certainly impermissible, there are some permissible maxims that involve not helping others (e.g., "never to help Nazis," "never to help slave-catchers," etc.). Now that maxim contradictories are no longer on the table, there seems to be very little reason to ascribe to an agent the maxim "sometimes to help others" simply because s/he has adopted the maxim "sometimes to help animal

rights activists." But even if such a claim is insisted upon, it is notable that the permissible maxims that involve not helping others have no plausible entailment relations with either the maxim "sometimes to help others" or the maxim "never to help others." Once again it follows that none of these maxims is obligatory.

However, the second line of attack is, I think, even more problematic for the proponent of the arguments I am arguing against. I think it is unlikely that any agent ever contemplates (in deliberation rather than in abstract consideration of Kant's or Kantian ethics) or adopts either the maxim "sometimes to help others" or the maxim "never to help others." The first is too vacuous to be considered as a principle of action; the second is too capacious for any agent realistically to contemplate. When using the notion of maxim contradictories, this did not seem problematic: maxim contradictories seem to be, of necessity, things that go on in the background. The problem with maxim contradictories is that too much is going on in the background, not that the maxims in question are artificial. However, when using the notion of maxim contraries and the deliberative field, the artificial nature of these maxims should raise eyebrows: if these maxims are not serious contenders for a deliberating agent, what is the sense in appealing to the deliberative field in arguing that an agent necessarily must adopt one of them? Very little. Or so I argue.

Conclusion

I have not claimed that Kant does not attempt to derive positive duties from the universalizability tests: my project in this paper was not a textual one. And I have not claimed that positive duties are superfluous, that any and all intuitions about them can be accounted

for by appeal to (for example) negative duties and general duties. If the arguments in this paper work, then I have shown that the standard account fails in its attempts to derive positive duties from the universalizability tests. But note that if the arguments in this paper work, then they show not merely that if the universalizability tests were one's only ethical standards, one never would be in a position to know whether any given maxim is obligatory. The idea is deeper: if the universalizability tests were one's only ethical standards, no given maxim would be obligatory. If the universalizability tests were one's only ethical standards, one would have sets of impermissible maxims and sets of permissible maxims, and it would be obligatory to adopt a permissible maxim. But there would be no sets of obligatory maxims. The universalizability tests do not give one the resources to say that an agent is acting badly unless one knows that the agent has adopted a maxim that is impermissible: if all that one knows is that there is some one maxim that an agent did *not* adopt, one does not have enough information to know whether the agent is acting badly if the universalizability tests are one's only ethical standards.

I would like to make one last point. Suppose I am correct and that it is impossible to derive positive duties from the universalizability tests. Would this show that these tests are pure formalisms, that if the universalizability tests were the only laws of morality, there could be no action purely from duty?[xlix] No. Although I do maintain that positive duties cannot be derived from the universalizability tests, I have said nothing in this paper about *negative* duties, and if it can be shown (as I believe it can) that negative duties can be derived from the universalizability tests, that might be enough to clear them of the charge of formalism.[l] But that, I think, is a project for another day.

[i] The positive/negative duty distinction is becoming more widespread in the secondary literature on Kant as the debate on this front attracts more interlocutors and as these interlocutors begin to articulate more narrowly defined arguments. See, for example, Allen Wood, *Kant's Ethical Thought* (Cambridge University Press: 1999), chapter 3 section 7.2 and Stephen Engstrom, *The Form of Practical Knowledge* (Harvard University Press: 2009), especially chapter 6 section 7 and chapter 7 sections 4-6. I shall be engaging with both Wood and Engstrom in more detail below.

The positive/negative distinction also appears in other debates, where it seems to have a similar import to the current one. See, for example, Özlem Ayse Özgür, "Human Rights Duties are Collective Duties of Justice," in Andreas Follesdal, Reidar Maliks (ed.s), *Kantian Theory and Human Rights* (Routledge: 2014): 89-112 (Özgür's definition of the distinction can be found at 109n4).

[ii] Engstrom argues that "positive duties can be brought to light through its [*viz.*, the Formula of Universal Law, one of the universalizability tests] application" (*The Form of Practical Knowledge*, *op. cit.*, 220). Engstrom wants to show (against people like me) that it is *not* the case that "the universalization test can perhaps show that certain maxims are permissible, but never, it seems, that a maxim is obligatory" (*ibid.*, 174); that it is *incorrect* to think that "the formula of universal law merely forbids acting on maxims that cannot be universalized and therefore…[is] incapable of generating any positive duties" (*ibid.*, 209). I shall engage with Engstrom's argument in more detail below.

[iii] First in part I at 4:402 and again in part II at 4:421.

[iv] Cf. J. Aule, "Aspekte des Universalisierungspostulats in Kants Ethik," *Neue Hefte für Philosophie*, vol. 22, 1983 and Thomas Pogge. "The Categorical Imperative," in Paul Guyer (ed.), *Kant's Groundwork of the Metaphysics of Morals: Critical Essays* (Rowman and Littlefield: 1998): 189-215, especially pp. 191-192.

[v] This is evident from the way in which Kant introduces FLN: "Since the universality of law in accordance with which effects take place constitutes what is properly called *nature* in the most general sense (as regards its form)—that is, the existence of things insofar as it is determined in accordance with universal laws—the universal imperative of duty also can go as follows…" (4:421).

[vi] If a maxim can be willed as a law of nature, it follows immediately that it can be conceived as such. So insofar as I am interested only in whether a maxim is permissible, it is not important for me whether the reason the maxim fails the test lies in a contradiction in conception or a contradiction in the will. (Referring here to what has come to be known as the distinction between the "contradiction in conception" test and the "contradiction in volition" or the "contradiction in the will" test. This terminology is sometimes attributed to O'Neill (see, for example, Allen Wood, *Kant's Ethical Thought*, *op. cit.*, p. 84).)

[vii] Henry Allison, *Idealism and Freedom* (Cambridge University Press: 1996), p. 160.

[viii] And, of course, elsewhere, including the *Metaphysics of Morals* (e.g., 6:446); his notes and fragments (e.g., R7264, 19:297); his lectures on ethics (e.g., 29:609); etc.

[ix] It will be noticed that whereas Allison contrasts narrow with *broad* duties, in the text of the *Groundwork for the Metaphysics of Morals* referred to here, narrow duties are contrasted with *wide* duties: "…the first is opposed to strict or narrower (unremitting) duty, the second only to wide (meritorious) duty…" (4:424). Figuring out which (if any) of these distinctions are equivalent is one of the textual challenges with which commentators grapple. However, it is not one that I shall enter into here. In any event, not much hinges on this for me; Allison himself puts off engagement in this complicated subject for the purposes of the paper I am discussing (see Henry Allison, *Idealism and Freedom*, *op. cit.*, 208n6), so I hope that I shall be excused for doing the same.

[x] See, for example, Mary Gregor, *Laws of Freedom* (Oxford University Press: 1963); Paul Eisenberg, "From the Forbidden to the Supererogatory: The Basic Ethical Categories in Kant's *Tugendlehre*," *American Philosophical Quarterly* vol. 3, No. 4 (Oct., 1966), 255-269; and Onora Nell (now O'Neill), *Acting on Principle* (Columbia University Press: 1975), especially chapter 4.

[xi] On p. 166 Allison focuses on the universalizability tests. Allison argues that Kant's attempt to ground the positive duty of beneficence on them in part II of the *Groundwork for the Metaphysics of Morals* fails. However, on p. 167 Allison notes that Kant thought that he could ground this duty on the so-called Formula of Humanity ("*So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means*" (4:429)). As Allison points out, according to Kant, if humanity is going to have its full effect on an agent, that agent must adopt others' ends as his/her own. It might be thought on these grounds that Allison takes the Formula of Humanity to be able to ground this duty even though the universalizability tests cannot. (Engstrom evidently reads Allison in this way: see Stephen Engstrom, *The Form of Practical Knowledge*, *op. cit.*, 174n9.) I would like to say two things about this.

First, Allison brings up Kant's views about the Formula of Humanity to make an *exegetical* point. In particular, Allison wants to show that *Kant* thought he could derive positive duties from the universalizability tests. Allison's argument is this: (1) Kant thought he could derive positive duties from the Formula of Humanity; (2) Kant thought the Formula of Humanity is equivalent to the universalizability tests; therefore (3) Kant might have thought that he could derive positive duties from the universalizability tests. So Allison is not bringing up the argument from the Formula of Humanity because he endorses it.

Second, there are good grounds for thinking that Allison does not endorse the argument from the Formula of Humanity. For one thing, Allison argues that "this notion of the conception of humanity as an end [in] itself as having its 'full effect'…is obscure and unanalyzed" (Henry Allison, *Idealism and Freedom*, *op. cit.*, p. 167). This suggests that he does not put much credence in the argument from the Formula of Humanity. For another thing, Allison's arguments in sections 2 and 3 (and perhaps most importantly, the argument in section 3 in which Allison takes himself to ground the claim that there are obligatory ends) make(s) no reference to the Formula of Humanity or to the universalizability tests. These arguments have to do only with the bare concept of a categorical imperative, the possibility of acting from duty and transcendental freedom. This suggests that Allison's argument about obligatory ends does not disambiguate between the different formulations of the Categorical Imperative. Finally, Allison concludes his discussion by suggesting that obligatory ends are inseparably connected to the first principle of the Doctrine of Virtue (the second half of the *Metaphysics of Morals*) rather than with anything in the *Groundwork for the Metaphysics of Morals*, and Allison says explicitly that this first principle "goes well beyond" that which is on offer in the *Groundwork for the Metaphysics of Morals* (Henry Allison, *Idealism and Freedom*, *op. cit.*, p. 168). This suggests that Allison does not take the Formula of Humanity (or anything else in the *Groundwork for the Metaphysics of Morals*) to be able to ground positive duties.

So it seems to me that it is reasonable to conclude that Allison is not confining himself to the universalizability tests in his paper (*pace* Engstrom), although it must be conceded that Allison does spend more time with the universalizability tests than with the other formulations.

[xii] See especially Henry Allison, *Idealism and Freedom*, *op. cit.*, section 3. Along the way (in section 2), Allison argues that Potter's earlier attempt at such a derivation fails. Potter's attempt hinges on the claim that if there is a categorical imperative, then it is possible to act from duty (Nelson Potter, "Kant on ends that are at the same time duties," *Pacific Philosophical Quarterly* vol. 66, Issue 1-2 (1985), pp. 78-92, see especially section 2). Allison argues that this focus on acting from duty is misguided. However, this seems somewhat disingenuous given that Allison replaces this focus with a focus on transcendental freedom and then goes on to claim that "the essential feature of transcendental freedom that is relevant to the present issue is that it involves not merely a causal independence, but also what, for want of a better term, I call 'motivational independence'" (*ibid.*, p. 162). However, as will become clearer below, Allison's and Potter's arguments are to some extent at cross-purposes with mine.

[xiii] At the level of exegesis, Wood argues that Kant did not think the various formulations of the Categorical Imperative on offer in the *Groundwork for a Metaphysics of Morals* are equivalent (Allen Wood, "The Supreme Principle of Morality," in Paul Guyer (ed.) *The Cambridge Companion to Kant and Modern Philosophy* (Cambridge University Press: 2006): 342-380). However, Engstrom argues that the various formulations are meant to be objectively equivalent (Stephen Engstrom, *The Form of Practical Knowledge, op. cit.*, especially chapter 4, section 2), and Pogge claims that "Kant wants these formulas to be equivalent" (Thomas Pogge, "The Categorical Imperative," *op. cit.*, p. 189).

At the level of philosophy, Korsgaard evidently does not think the various formulations are equivalent, for she takes them to render different verdicts about the permissibility of some maxims (Christine Korsgaard, *Creating the Kingdom of Ends* (Cambridge University Press: 1996), chapter 5). O'Neill, by way of contrast, argues that the various formulations are, at bottom, the same (Onora O'Neill, *Constructions of Reason* (Cambridge University Press: 1989), chapter 7). This debate is especially complicated by virtue of the fact that even when two commentators agree that the formulations are (or are not) philosophically equivalent, they might have very different ideas about why. For example, Wood evidently thinks that the formulations are not equivalent, but his reasons for doing so are very different from Korsgaard's (Allen Wood, *Kant's Ethical Thought*, *op. cit.*, chapters 3-4). Similarly, Melnick seems to regard the various formulations as equivalent, but his interpretations of them differ markedly from O'Neill's (Arthur Melnick, "Kant's Formulations of the Categorical Imperative," *Kant studien* vol. 93 (2002), pp. 291-308).

Complicating matters still more, both of these debates (the exegetical and the philosophical debates) intersect with debates about how many formulations of the Categorical Imperative there are in the first place. For some representative publications in this last line of inquiry consider, for example, Chroust, who takes there to be three and then discovers a fourth at the foundation of Kant's theory of criminal justice (Anton-Hermann Chroust, "About a fourth formula of the Categorical Imperative of Kant," *The Philosophical Review*, Vol. 51, No. 6 (Nov., 1942), pp. 600-605). Alternatively, consider Paton, who takes there to be five formulations but does not include Chroust's fourth formulation on his list (H.J. Paton, *The Categorical Imperative*, *op. cit.*, p. 129). Paton's work on this front sparked a number of responses. See, for example, Baker and, subsequently, Nuyen (Judith Baker, "Counting Categorical Imperatives," *Kant studien* vol. 79 (1988), pp. 389-406; A. T. Nuyen, "Counting the Formulas of the Categorical Imperative: One Plus Three Makes Four," *History of Philosophy Quarterly*, Vol. 10, No. 1 (Jan., 1993), pp. 37-48).

[xiv] Korsgaard sums up some of the earlier debates about the various interpretations on offer in *Creating the Kingdom of Ends, op. cit.*, chapter 3. The discussion has continued since that time (see, e.g., Shawn D. Kaplan, "A critique of the practical contradiction for testing maxims," *Kantian review*, Vol. 10, No. 1 (2005), pp. 112-127). But again: I shall not be taking a stand on these issues here.

[xv] Aune argues that FUL and FLN are different in at least one respect: unlike FLN, imperfectly rational beings (like humans) cannot apply FUL to experience in any direct way (*Kant's Theory of Morals*, *op. cit.*, chapter 2 section 3). Galvin disagrees: he argues that there is no significant difference between FUL and FLN in regard to their application to experience (Richard Galvin, "The universal law formulas," in Thomas Hill, Jr. (ed.) *The Blackwell Guide to Kant's Ethics* (Wiley-Blackwell: 2009): 52-82. For the relevant claims, see pp. 58-59). I want to be clear that I am not taking a stand here on the equivalence (or nonequivalence) of FUL and FLN. I say only that Kant seems to regard them as equivalent in the sense noted above and that there is no need to distinguish between them for the purposes of this paper.

33

[xvi] In this I take myself to be following most interpreters. See, for example, H.J. Paton, *The Categorical Imperative* (Hutchinson's University Press: 1946), p. 141:

> It is quite clear that on Kant's view to act on maxims which will not meet the requirements of universality is to act wrongly. It is also clear that it is permissible…to act on maxims which meet this requirement.

See also Stephen Engstrom, *The Form of Practical Knowledge*, *op cit.*, chapter 7 section 6 and Thomas W. Pogge, "The Categorical Imperative," in Paul Guyer (ed.), *Kant's Groundwork of the Metaphysics of Morals: Critical Essays* (Rowman and Littlefield: 1998), pp. 189-213, and especially pp. 189-190, where, in a discussion of universalizability, Pogge remarks that "[t]he categorical imperative is postulated as a *necessary and sufficient* condition for the *permissibility* of maxims."

Aune seems to think that on a strict reading, only the first conditional in the text above (an agent acts on a maxim that cannot be willed at the same time as a universal law, then that agent is acting impermissibly) is contained in FUL. However, he says that on a looser reading, both conditionals are contained in FUL, and because he takes it to be clear that Kant is committed to both conditionals, he does not think much follows from this (Bruce Aune, *Kant's Theory of Morals* (Princeton University Press: 2014), chapter 2 section 1).

I should point out that if one were to contend that only the first conditional follows from FUL, this only would strengthen my argument. So although I do think that on a charitable interpretation of FUL, the biconditional can be read into Kant, if I am wrong about this, it will not impugn the argument in this paper—all the more especially because in this paper I am interested in addressing modern derivations of positive duties from FUL rather than in engaging in textual disputes, and these modern derivations all appeal to FUL as a biconditional. Engstrom, mentioned at the beginning of this footnote and about whom I shall say more later, is a conspicuous example of this. After setting forth his derivation of positive duties from the universalization tests, he argues as follows:

> …absolutely and originally, then, the moral law lies in the positive direction to act according to maxims that can be willed as universal laws; the prohibition against acting according to maxims that cannot be so willed…[derives] from the original positive direction through the latter's relation to imperfect particular wills." (Stephen Engstrom, *The Form of Practical Knowledge*, *op. cit.*, pp. 222-223)

[xvii] A notable exception to the tendency to view the universalizability tests as permissibility tests can be found in Patricia Kitcher, "Kant's Argument for the Categorical Imperative," *Noûs* vol. 38, No. 4 (2004), pp. 555-584. Kitcher argues that FUL "…presupposes that all actions are either obligatory or forbidden… that there are no morally acceptable, but merely permissible, actions" (p. 572). The word 'presupposes' here is important: Kitcher concedes that Kant "*intended* FUL to be a test of the permissibility of actions" (*ibid.*, my emphasis); her point is that premises in the derivation of FUL in the *Groundwork for the Metaphysics of Morals* reveal that it cannot serve this purpose.

Kitcher's argument for this point is based on two key ideas. First, she claims that on Kant's account, part of the concept of a categorical imperative is that an agent must be able to derive his/her maxims from an imperative. Second, she maintains that every imperative is "an 'ought' or an 'ought not'" (*ibid.*). Kitcher takes this to be "fatal" to FUL, and she proposes a workaround that "is deeply Kantian in spirit" and that enables FUL to distinguish between obligatory, permissible and forbidden maxims (*ibid.*). I would like to make four quick points about this.

First, I have some misgivings about Kitcher's reconstruction. She claims that in section I of the *Groundwork for a Metaphysics of Morals*, "Kant has already argued…for the positive conclusion that the morality of an action depends on its motivation" (Patricia Kitcher, "Kant's Argument for the Categorical Imperative," *op cit.*, p. 565). Given that Kitcher thinks that "…the only two varieties [of motives] Kant allows are 'duty' and 'self-love'…[and] maxims of self-love, as such, have no moral quality" (*ibid.*), it follows that an action is moral if and only if it is done from duty. But this risks a *reductio*: if an action that is not from duty is not moral, then an action that is performed *from* duty is not a duty to begin with. Kant's introduction of actions that are "*in conformity with duty* but not *from duty*" at 4:398 (in section I) seems to get around this in a way that is blocked to Kitcher (namely, by severing the tie between the morality of an action and its motive).

Second, I am skeptical of Kitcher's attribution to Kant of the claim that part of the concept of a categorical imperative is that an agent must be able to derive his/her maxims from an imperative. She bases this attribution on the following passage:

> …when I think of a *categorical* imperative, I know at once what it contains. For, since the imperative contains, beyond the law, only the necessity that the maxim be in conformity with this law, while the law contains no condition to which it would be limited, nothing is left with which the maxim is to conform but the universality of a law as such; and this conformity alone is what the imperative properly represents as necessary. (4:420-421; Kitcher cites this passage on p. 555)

According to Kitcher, the word 'conform' should be interpreted as "derivable": "…I take his [*viz.*, Kant's] view to be that…[a maxim] 'conforms' to an imperative just in case the agent can derive it from that imperative" (Patricia Kitcher, "Kant's Argument for the Categorical Imperative," *op cit.*, p. 566). But note that this would mean Kant is saying in the second half of this passage that a maxim must be derivable from the universality of law as such, which does not work: the maxim will have content (matter)—Kant is saying that in addition to this content, it must have a certain form if it is to be consistent with a categorical imperative.

Third, the fact (if it is one) that every imperative is "an 'ought' or an 'ought not'" does not suffice to rule out the category of morally permissible actions. This is for the simple reason that the imperative could be disjunctive. For example, the command could be something like, "you ought to X or to Y." To make this more concrete, someone might say, "you ought to serve on at least *one* of these committees: the Academic Affairs committee, the library committee, or the Academic Planning committee." In such a case, the "mere" permissibility of serving on the library committee rather than the Academic Planning committee is far from nugatory.

Fourth and finally, Kitcher's ultimate workaround will turn out to fail for exactly the same reason the standard account of how to derive positive duties from universalizability tests fails: Kitcher's workaround turns out to be committed to an objectionable form of maxim realism. I shall have more to say about this below.

[xviii] Paul Guyer, *Kant's "Groundwork for the Metaphysics of Morals": A Reader's Guide* (Continuum Press: 2007), p. 143.

35

[xix] Barbara Herman, *The Practice of Moral Judgment* (Harvard University Press: 1993), p. 63. Baron and Fahmy also might buy into this when they argue that Kant's ethics "tells us…that a particular maxim is, or is not, permissible (or that it is, or is not, required)" (Marcia Baron and Melissa Fahmy, "Beneficence and Other Duties of Love in *The Metaphysics of Morals*," in Thomas Hill, Jr. (ed.) *The Blackwell Guide to Kant's Ethics* (Wiley-Blackwell: 2009): 205-222, p. 206). However, there is no way to be sure: they do not explain how they take Kant's ethics to deliver these required maxims.

[xx] Christine Korsgaard, *Creating the Kingdom of Ends*, *op. cit.*, p. 63. Paton also uses "opposite": "…nevertheless Kant manages to extract from this formula [*viz.*, FUL] positive duties, like the duty of developing talents and of helping others…Kant at least professes to establish these duties by showing that to will the opposite is to fall into self-contradiction" (*The Categorical Imperative*, *op. cit.*, 143). Paton seems to be engaged primarily in exegesis in this passage: by talking about what Kant professes, Paton perhaps wants to avoid taking a stand on whether such a strategy works. As pointed out above, my goal in this paper is not to show that this strategy is not exegetically warranted. However, it is worth noting that at least in *this* instance, the strategy is *not* exegetically warranted. Paton refers to the following passage from the *Groundwork for the Metaphysics of Morals* to substantiate his reading:

> If we now attend to ourselves in any transgression of a duty, we find that we do not really will that our maxim should become a universal law, since that is impossible for us, but that the opposite of our maxim should instead remain a universal law, only we take the liberty of making an exception to it for ourselves (or just for this once) to the advantage of our inclination. (4:424)

But a close inspection of this passage reveals that Kant is not advocating the strategy of deriving positive duties by showing that to will the opposite is to fall into self-contradiction. Rather, in this passage Kant is giving a schema for moral transgressions: they involve making an exception to a rule that one *as a matter of fact* (rather than as a matter of obligation) wills to remain a universal law.

[xxi] *Ibid.*, pp. 63-64.

[xxii] See note 39 (below) for evidence that commentators who derive positive duties by appeal to maxim contraries and the deliberative field also take the positive/negative duty distinction is to cut across the perfect/imperfect duty distinction.

[xxiii] Jonathan Harrison, "Kant's Examples of the First Formulation of the Categorical Imperative," *Philosophical Quarterly* vol. 7, No. 26 (Jan., 1957), pp. 50-62. The text above is taken from p. 52. In the preamble to a discussion of the duty of beneficence, Stohr uses 'negation' rather than 'contradictory', but the argument is otherwise almost identical to Harrison's:

> If the maxim [to which the first formulation of the Categorical is applied] fails the universalizability test, the action is forbidden. An action is required if the negation of its maxim fails the test. Maxims that otherwise pass the test are permissible. (Karen Stohr, "Kantian Beneficence and the Problem of Obligatory Aid," *Journal of Moral Philosophy* vol. 8 (2011), pp. 45-67. This excerpt is taken from p. 50.)

However, there is some ambiguity here in that Stohr says that an *action* is required if the negation of its *maxim* fails the test. Given that she goes on in the next paragraph to suggest that "[i]mperfect duties…are not duties to perform or refrain from particular actions, but rather to adopt maxims," I suggest that she is saying that when the negation of a maxim fails the universalizability tests, then there is a positive duty, and this duty is a perfect duty (*ibid.*). If my suggestion is correct, then this is more evidence that the positive/negative distinction cuts across the perfect/imperfect distinction. Indeed, at the beginning of the paragraph from which the block quotation above is excerpted, Stohr claims that "perfect duties prescribe or prohibit specific actions whereas imperfect duties are duties to adopt a maxim" (*ibid.*). However, even if my suggestion is incorrect, it should be noted that Stohr is buying into the standard derivation of positive duties in this passage. I note in passing that in this excerpt, Stohr seems to be assuming that there is either a one-one or a one-many relationship between maxims and actions. I suspect that this assumption is false, but see Onora Nell, *Acting on Principle*, *op. cit.*, chapter 2, especially pp. 62-63.

[xxiv] Philip Stratton-Lake, *Kant, Duty, and Moral Worth* (Routledge: 2000), p. 50.

[xxv] Some tightening here is needed in order to account for the possibility of changing maxims through time (maybe at time t_1 agent A would tell a lying promise, but at time t_2 (perhaps after having seen, firsthand, how harmful lying promises can be to others) A would not; at some time between t_1 and t_2, A converted). I think it would suffice to add the rider "at the same time" to the definition of maxim contradictories given above. But hopefully the main idea here is clear; as will emerge below, this point about conversion does not have much bearing on my argument.

[xxvi] If it turns out to be the case that both O_m and C_m pass the universalizability tests, then, of course, there is no positive duty. But that is no embarrassment to the proponent of this account. In fact, quite the alternative: such a case might be deemed helpful because both the maxims presumably would be "merely" permissible and therefore would enable the account to avoid rigorism. An example is easy to construct using the beneficence maxim as a model: presumably the maxim "never to eat fish" passes the universalizability tests, and presumably its contradictory, "sometimes to eat fish," also passes the universalizability tests. If this is so, proponents of this account probably would say that it is permissible for agents to adopt either of these maxims. (I suppose some might argue that a strict vegetarian diet is obligatory. I am not going to dispute that; hopefully the point is clear even if the particular example used to illustrate it, which is loosely based on some remarks made by Kant (6:409), is unpersuasive.)

If it turns out to be the case that both O_m and C_m fail the universalizability tests, things are a bit more complicated. Because (again, by definition of a maxim contradictory) all agents adopt either O_m or C_m, if there are such maxims, then all agents are always behaving impermissibly. I suppose that the proponent of this account might assert that such a case never will arise. Alternatively, s/he might assert that in such a case, both maxims are permissible. To say that both maxims are permissible would be a revision of Kant: I am unable to find any place in the texts where he suggests that a nonuniversalizable maxim can be permissible. But many commentators would be fine with that.

O'Neill grapples with these issues at some length in the context of a derivation of positive duties using maxim contraries (rather than maxim contradictories). Here is her conclusion:

> To go through this procedure is to check whether a maxim and its contrary can be universalised without contradiction. If the procedure shows that both can be universalised or that neither can be, acts conforming to either maxim or to its contrary are merely permissible. If only the maxim cannot be universalised, acts conforming to it are forbidden; if only the contrary cannot be universalised, acts conforming to the maxim are obligatory. (Onora Nell (O'Neill), *Acting on Principle*, *op. cit.*, p. 166.)

O'Neill argues that neither "nonreciprocal action" maxims (e.g., "I will buy clockwork trains but not sell them") nor their contraries (in this case, "I will sell clockwork trains but not buy them") are universalizable, so they are permissible (*ibid.*). However, there are three things that are notable about this.

First, O'Neill does not argue that these maxims are permissible on the grounds cited just above (namely, that otherwise this would entail the counterintuitive result that all agents would be behaving impermissibly at all times). The grounds cited above do not hold here because she is dealing with contraries rather than contradictories. Rather, O'Neill's argument is that to classify these maxims any other way "would not have been intuitively acceptable" (*ibid.*). She admits that "this move is…a shift from exegesis to reconstruction" (*ibid.*).

Second, O'Neill's claims about nonreciprocal action maxims are somewhat problematic. Consider, for example, "to skin but not to be skinned" or "to eviscerate but not to be eviscerated" and their corresponding contraries. At least some of these four maxims seem (to my mind) intuitively impermissible. But O'Neill must conclude that they are all permissible. I am not sure what part of her account O'Neill would be most ready to give up, but I do think this shows she has to give up something.

Third, note that part of the problem is that O'Neill takes maxim contraries to come in pairs. This assumption is treated at some length in the next section of this paper. But consider the following three maxims: (1) "never to help anyone," (2) "always to help the Nazis catch Jews" and (3) "always to help Jews escape from the Nazis." These are presumably maxim contraries, and presumably O'Neill would argue that neither (1) nor (2) is universalizable whereas (3) is. This then generates a contradiction: the fact that neither (1) nor (2) is universalizable entails (by O'Neill's lights) that (1) is permissible; but the fact that (1) is not universalizable whereas (3) is universalizable entails (again, by O'Neill's lights) that (1) is impermissible; so (1) is both permissible and impermissible. One also can generate a contradiction using maxims (1) and (3) from above and (4) "always to help runaway slaves escape from slave catchers and never to help the Nazis catch Jews." A chain of reasoning similar to the one just rehearsed will show that (4) is both obligatory and *merely* permissible.

38

xxvii According to Yudanin, "given certain worldviews, the maxim of non-beneficence can be universalized without contradiction, and the maxim of beneficence will be irrational" (Michael Yudanin, "Can positive duties be derived from Kant's Categorical Imperative?" *Ethical Theory and Moral Practice* vol. 18, No. 3 (June, 2015), pp. 595-614, p. 605). Thus, Yudanin objects to the derivation of positive duties at exactly this point in my reconstruction of it: he argues that although *some* agents will not be able to universalize the maxim "never to help others," *some* will (depending on their respective worldviews).

To illustrate this idea, Yudanin argues that "giving help is contradictory within the particular worldview" of a social darwinist who is committed to "survival of the fittest" as a law of nature (Michael Yudanin, "Can positive duties be derived from Kant's Categorical Imperative?" *op. cit.*, p. 604). He infers from this that a social darwinist would be obligated to adopt a maxim "never to help" rather than a maxim "to help some others sometimes." Thus, Yudanin concludes, there is no positive duty to adopt the maxim "to help some others sometimes." Because he contends that this problem about worldviews infects all attempts to derive positive duties from FUL (not merely attempts to derive the duty of beneficence), Yudanin asserts that no positive duties can be derived therefrom. Moreover, he argues that this problem "is not specific to this particular formulation of the Categorical Imperative [*viz.*, FUL]—on the contrary, it is a principal issue that touches the core of Kant's moral theory" (Michael Yudanin, "Can positive duties be derived from Kant's Categorical Imperative?" *op. cit.*, p. 607). I would like to say three things about Yudanin's line of argument.

First, Yudanin's objection presupposes a contentious interpretation of the Categorical Imperative. In particular, Yudanin evidently is committed to the view that (1) the maxims that *I* can will at the same time as universal laws might be different from those that an ideal agent can will at the same time as universal laws and (2) if there is such a difference, the maxims that are permissible for me are those that *I* can will at the same time as universal laws. It is true that not all commentators are clear about where they stand on (1) and (2). But some are. For example, in Rawls' discussion of the duty of beneficence, he is quite explicit about the fact that various idealizations are at work and therefore that agents' opposing worldviews do not come into play (John Rawls, *Lectures on the History of Moral Philosophy* (Harvard University Press: 2000), especially pp. 172-176; see also John Rawls, "Themes in Kant's Moral Philosophy," in Eckart Förster (ed.), *Kant's Transcendental Deductions* (Stanford University Press: 1989), pp. 81-113). For this reason, I am somewhat skeptical about whether many proponents of positive duties would concede that Yudanin's argument makes contact. This point is exacerbated by the fact that Yudanin does not seem to have an argument for his interpretation of the Categorical Imperative.

Second, Yudanin's definition of "positive duty" is different from mine. As may be inferred from the nature of his objection, Yudanin thinks that a positive duty can be derived from the universalizability tests if but only if it can be shown that there is a maxim that *all* agents are obligated to adopt. By way of contrast, I maintain that a positive duty can be derived from the universalizability tests if but only if it can be shown that there is a maxim that *some* agent(s) is (are) obligated to adopt. From this it may be seen that my thesis about the underivability of positive duties entails Yudanin's, but the reverse does not hold. Indeed, even Yudanin's formulation of his objection (reproduced in the first sentence of this footnote) suggests that positive duties in *my* sense of the term can be derived from the universalizability tests: Yudanin's point is that these positive duties do not hold for all agents. Thus, even if Yudanin's objection is granted, the fact that he concedes that *some* agents would be obligated to will the maxim of beneficence puts him on the other side of the aisle from me, so to speak.

Third and finally, it is perhaps worth pointing out that if Yudanin's interpretation of the Categorical Imperative is accepted along with his presuppositions about the universal scope of duties, then it looks like *no* duties will be derivable from the Categorical Imperative, for differing worldviews will infect the derivations of negative duties just as much as positive duties. Yudanin initially denies this (Michael Yudanin, "Can positive duties be derived from Kant's Categorical Imperative?" *op. cit.*, p. 604). But by the end of the article, he is ready to concede that what would be permissible for an agent with a worldview from a few centuries ago might not be permissible for an agent today (Michael Yudanin, "Can positive duties be derived from Kant's Categorical Imperative?" *op. cit.*, p. 612). Thus, Yudanin's line of attack seems to be at cross-purposes with the dialogue about positive duties.

xxviii Herman argues that the duty is never to make a deceitful promise (Barbara Herman, *The Practice of Moral Judgment*, *op. cit.*, p. 171)—and in a footnote she refines this further to a duty never to make a deceitful promise from self interest (171n16). However, as will emerge below, these details are not important for my purposes.

xxix For a helpful discussion along these lines, see Allen Wood, "Kant on duties regarding Nonrational Nature," *The Aristotelian Society Supplementary Volume* vol. 72, No. 1 (June, 1998), pp. 189-210 or Allen Wood, *Kantian ethics* (Cambridge University Press: 2008), especially chapter 5 section 4.

xxx I am purposefully confining myself to Kantian (rather than Kant's) ethics here: as above, my goals are primarily philosophical rather than exegetical. However, it is worth pointing out that according to Kain, on Kant's account moral status begins at conception (Patrick Kain, "Kant's defense of Human Moral Status," *Journal of the History of Philosophy* vol. 47, No. 1 (Jan. 2009), pp. 59-101). The first part of Kain's argument for this hinges on a parallel between Kant's biological and psychological commitments: Kant is committed to the claim that at conception, a new organism is created with various original biological predispositions (which ones it has depends on the (fixed) species to which it belongs) and, simultaneously, this organism is fused with a soul that has various original psychological predispositions (*ibid.*, section 2, especially p. 82). The second part of Kain's argument hinges on a claim about moral status: (in humans) these psychological predispositions include the predisposition to personality, and Kant ties this particular predisposition to moral status (*ibid.*, section 3, especially p. 88). Putting these two parts of the argument together, it follows that (for Kant) moral status begins at conception. I would like to say three things about this argument.

First, I am skeptical of the second part of this argument and, in particular, of the claim that Kant ties the predisposition to personality to moral status (or, alternatively, that for Kant "[w]hat qualifies us for that special dignity as ends-in-ourselves…is the predisposition to personality" (*ibid.*, p. 88)). Kain cites five main texts to substantiate this claim: 4:440, 5:86-88, 5:161-163, 6:28 and 6:49. For the sake of brevity, I shall confine myself to these five texts, beginning with the two from the *Religion within the Boundaries of Mere Reason.* To be fair, at 6:49 Kant does say that it is the "original moral predisposition" for which "admiration is legitimate and uplifting as well." But at 6:28, Kant does not make any claims about moral status. However, he does distinguish in this passage between the predisposition to personality and personality itself: "[t]he idea of the moral law alone, together with the respect that is inseparable from it, properly cannot be called a *predisposition to personality*; it is personality itself (the idea of humanity considered wholly intellectually)" (6:28). This is relevant because in both of the passages Kain cites from the *Critique of Practical Reason,* Kant suggests that it is personality rather than the predisposition to personality that is linked to moral status: Kant tells us that it is "their *personality*, by which alone they are ends in themselves" (5:87) and that reflection on the moral law within me "infinitely raises my worth as an *intelligence* by my personality" (5:162). (Indeed, Kant goes on at 5:163 to exhort his readers to develop their "moral predispositions," showing that here, too, he distinguishes between the predisposition and personality itself.) At 4:440, Kant does not mention the predisposition to personality. However, he does suggest that "…the dignity of humanity consists just in this capacity to give universal law," which (to my eye) looks more like personality than the predisposition thereto. So I think that the evidence for the second part of Kain's argument is ambivalent at best, and this is a serious problem for his thesis: if it is personality rather than the predisposition to personality that is tied to moral status, then moral status might not begin at conception even if the predisposition to personality does.

Second (and overlooking these textual skirmishes), provided that these souls are created (i.e., temporally infinite in at most one direction, the forward one) the point made in the main text still stands: upon creation, these souls acquire an infinite number of maxims.

Third and finally, even if the souls are infinite in both directions, the deeper point that the argument in this section hinges on remains: if there are these maxim-contradictory pairs, then these souls will have many maxims that they do not possess the concepts for and that are in no way, shape or form action guiding.

xxxi The actual age at which this happens is irrelevant. I have chosen 18-21 because that is when kids transition into adulthood in the eyes of the law.

xxxii For a helpful discussion of this point, see Nelson Potter, "Maxims in Kant's Moral Philosophy," *Philosophia* vol. 23, No. 1-4 (1994): 59-90, especially section 5.

xxxiii For a helpful discussion of this distinction see, e.g., Henry Allison, *Kant's Theory of Freedom* (Cambridge University Press: 1990), chapter 5 section 1.

xxxiv O'Neill, an advocate of the position that specific intentions do not count as maxims, illustrates the distinction by contrasting the underlying intention "to make a visitor welcome" with the more specific intentions "to warm the tea pot" and "to offer the sugar" (Onora O'Neill, "Kant after Virtue," *Inquiry* vol. 26, No. 4 (1983), pp. 387-405, especially p. 394). Hilly Flitcraft is made good use of by Allen Wood, *Kant's Ethical Thought*, *op. cit.*, pp. 102-106. Wood even includes him in the index.

xxxv As noted by Allison, "…as important as the distinction may be for Kant's moral philosophy, the line between…underlying and specific intention, is not always an easy one to draw" (Henry Allison, *Kant's Theory of Freedom*, *op. cit.*, p. 94.

xxxvi O'Neill claims that Kant often proposes such reasoning. However, she does not provide any citations to substantiate this claim, and she goes on to concede that "such tests are not decisive when they appeal to counterfactual possibilities, given that the consciousness of agents is not transparent" (Onora O'Neill, "Kant After Virtue," *op. cit.*, p. 395).

xxxvii Cf. Henry Allison, *Kant's Theory of Freedom*, *op. cit.*, p. 90:

> …since maxims are self-imposed rules, one cannot make something one's maxim without in some sense being aware of it as such, or at least without the capacity to become aware of it…This does not entail…that we must explicitly formulate our maxims to ourselves before acting. The point is rather a conceptual one: namely, that I cannot act on a principle…without an awareness of that principle, although I need not be explicitly aware of myself *as* acting on that principle. Moreover, it must be possible in subsequent reflection to discover and articulate (albeit not in an indefeasible way) the maxims on which one acts…

xxxviii Recall from note 17 that according to Kitcher, FUL faces a problem because it presupposes that all actions are obligatory or forbidden: there is no room for the merely permissible (Patricia Kitcher, "Kant's Argument for the Categorical Imperative," *op. cit.*, section 6). Kitcher then proposes a solution to this problem: a workaround whereby FUL can be used to derive obligatory, permissible and forbidden maxims.

        Kitcher argues for all of this based on (1) her reconstruction of the derivation of FUL and (2) a claim about imperatives. I raised problems for both (1) and (2) in note 14, but I also promised to explain why Kitcher's ultimate workaround commits her to maxim realism. Now I shall try to make good on that.

        Unlike the standard account I attacked in this section, Kitcher does not rely on the notion of a maxim contradictory. Instead, she argues that if the maxim has the form "I will…" and it passes FUL, then it is obligatory, and if a maxim has the form "I will…if I like" and it passes FUL, then it is "merely" permissible:

> An agent can derive the maxim, 'I *to do* ACP' only if he adopts (however temporarily) the law, 'Everyone ought to do ACP'…An agent can derive his maxim, 'I to do ACP, if I like' form a law of permission, 'Everyone to do ACP, if he likes' only if he adopts that law of permission. (Patricia Kitcher, "Kant's Argument for the Categorical Imperative," *op. cit.*, p. 576)

The reason this still commits Kitcher to an objectionable form of maxim realism is that on this account there still (notwithstanding the fact that there is no appeal to maxim contradictories) will be a list of infinitely many maxims associated with each agent: all of the maxims that have the form "I will…" and that pass the universalizability tests.

        Kitcher might object that this is not quite right: strictly speaking, all of the maxims of the form "I will…" that pass the universalizability tests will be *obligatory*. This is not quite the same thing as saying that all agents will have them. So she can avoid objectionable maxim realism, after all.

        I would agree: Kitcher can avoid objectionable maxim realism. But the price is that all agents that do not have these infinitely many maxims associated with them will be behaving impermissibly. To put it plainly, Kitcher is between a rock and a hard place. The basic problem here, as with the standard account above, is with the kind of obligation that is being derived: timeless, obligatory principles the deontic status of which has nothing to do with the circumstances in which an agent finds him/herself. The argument I confront in the next section gets over this problem, but as I shall argue, in so doing, it renders it impossible to derive positive duties from the universalizability tests.

xxxix Paul Guyer, *Kant* (Routledge: 2006), p. 194. It should be noted that the book from which this passage is excerpted was published a year before the book from which the passage reproduced at the beginning of the previous section of this paper was excerpted. O'Neill uses the same line of reasoning in *Acting on Principle* as Guyer uses in the passage quoted above. As evidence for this, consider the following passage:

> In the case that a maxim can be consistently universalized, a second application of the contradiction in conception test to the maxim's contrary is needed to determine the deontic status of acts conforming to the maxim. If an agent's maxim is a specification of (3) [one of a series of maxim "forms" specified earlier in the book], of form C. To do A if B, and no contradiction emerges when C is universalized, then the same test must be applied to D. To omit A if B. If the test now yields a contradiction, then it is forbidden to omit A if B. If no contradiction emerges from either the test of C or the test of D then it is merely permissible to do A if B. (Onora Nell (O'Neill), *Acting on Principle*, *op. cit.*, pp. 162-163)

Note the reference to the contradiction in conception test in this passage: this passage is intentionally meant to be about paradigmatically *perfect* duties, like the duty to keep one's promises. Thus it may be seen that it is not only proponents of maxim contradictories who advocate for positive perfect duties.

[xl] Christine Korsgaard, *Creating the Kingdom of Ends*, *op. cit.*, pp. 152-153. It should be noted that the quotation used in the previous section of this paper came from a different chapter of *Creating the Kingdom of Ends*. Both chapters were published as separate articles (in different years) prior to being published in *Creating the Kingdom of Ends*. See Christine Korsgaard, "Kant's Analysis of Obligation," *The Monist* vol. 72, No. 3 (1989): 311-340 and Christine Korsgaard, "The right to lie: Kant on dealing with evil," *Philosophy and Public Affairs* vol. 15, No. 4 (1986): 325-349.

In *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*, Korsgaard seems to be committed to the view that positive duties can be derived from the universalizability tests. But it is unclear to me whether she appeals to something like the maxim contradictory account or something like the maxim contrary account. For example, on page 184 she argues as follows:

> Using this formulation [*viz.*, *"I ought never to act except in such a way that I could also will that my maxim should become a universal law"*] we discover that a maxim must be regarded as a law by discovering that the opposite maxim—the maxim of not doing the action in question—cannot be regarded as a law. (Christine Korsgaard, *The Constitution of Agency: Essays on Practical Reason and Moral Psychology* (Oxford University Press: 2008), p. 184.)

Korsgaard's use of the word 'discover' might be an appeal to the deliberative field. On p. 194 Korsgaard argues that "[u]nder the Formula of Universal Law you arrive at the duty of helping when you consider the maxim of not helping." Again, the use of the second person and the word 'consider' might be an appeal to the deliberative field.

In *Self-Constitution: Agency, Identity, and Integrity*, Korsgaard seems to endorse a maxim contradictory account:

> A maxim is fit to be a law in one sense, the sense corresponding to permissibility, if it could function as a law. It is fit to be a law in a stronger sense, the sense corresponding to obligation, if it not only can but must be a law. The way we ascertain this is by showing that the maxim of doing the opposite is unfit to be law, and must be rejected. So, for instance, when we find that we cannot will the maxim of false promising just to get a little ready cash, we arrive at the duty of promising sincerely for purposes of this kind. (Christine Korsgaard, *Self-Constitution: Agency, Identity, and Integrity* (Oxford University Press: 2009), p. 16.)

I am unsure whether Korsgaard ultimately advocates a maxim contradictory derivation or a maxim contrary/deliberative field derivation of positive duties. Perhaps she is of two minds. But one thing does seem to me to be clear: Korsgaard advocates the claim that positive duties can be derived from the universalizability tests throughout her work.

[xli] Stephen Engstrom, *The Form of Practical Knowledge*, *op. cit.*, p. 211. It is perhaps worth pointing out that when Engstrom discusses positive duties earlier in his book, he cites Wood as someone who thinks that no positive duties can be derived from the universalizability tests (Stephen Engstrom, *The Form of Practical Knowledge*, *op. cit.*, 174n9). Wood certainly does believe this now (personal communication). However, in the passage to which Engstrom refers us, Wood is arguing that the universalizability tests *can* ground a positive duty (namely, beneficence) even if this is "the only…positive duty derivable from it [FLN]" (Allen Wood, *Kant's Ethical Thought*, *op. cit.*, 359n26).

[xlii] Stephen Engstrom, *The Form of Practical Knowledge*, *op. cit.*, pp. 216-217.

[xliii] See, for example, the following passage from *The Practice of Moral Judgment*: "…insofar as Kantian duties necessitate, they do not do so directly. Necessitation (obligation) is always the outcome of deliberation" (Barbara Herman, *The Practice of Moral Judgment*, *op. cit.*, p. 172). I should point out that this quotation comes from chapter 8, "Obligation and Performance." The quotation reproduced in the previous section was taken from chapter 3, "Mutual Aid and Respect for Persons," which initially was published as an article in 1984 (Barbara Herman, "Mutual Aid and Respect for Persons, *Ethics* vol. 94, No. 4 (1984), pp. 577-602).

[xliv] Herman defines the concept of the deliberative field in a number of places. For example, consider the following passage: "I use the device 'deliberative field' to represent the space in which an agent's rational deliberation takes place. It is constructed by the principles and commitments that express her values" (Barbara Herman, "Making room for character," in Stephen Engstrom and Jennifer Whiting (eds.), *Aristotle, Kant and the Stoics: Rethinking Happiness and Duty* (Cambridge University Press: 1996), pp. 36-60, p. 49). Alternatively, consider this passage: "[a]n agent's deliberative field is the normative space constructed by the principles she accepts—usually an ordered array of moral and nonmoral principles" (Barbara Herman, *Moral Literacy* (Harvard University Press: 2007), pp. 40-41).

[xlv] This line of argument is formulated explicitly in *Acting on Principle*:

> It should be noted that this argument [that is, an argument to show that the maxim "to neglect everything needed to help the needy," which is labeled "1" on p. 182, cannot be willed as a universal law] shows only that a maxim of giving *no* help to *any* other in need is contrary to duty…Complete lack of beneficence is contrary to duty, but we do not have a duty to help all who need help…Duty requires only the adoption of the contrary of 1, "I will do some of what is needed to help the needy." (Onora Nell (O'Neill), *Acting on Principle*, *op. cit.*, p. 183)

It should be noted that despite the use of the word "contrary," O'Neill comes close in this passage (and others) to the kind of derivation used by proponents of maxim contradictories.

[xlvi] Stephen Engstrom, *The Form of Practical Knowledge*, *op. cit.*, chapter 3, especially section 4.

[xlvii] I am not persuaded that this assumption is true. But consider the following passage from *Kant's Theory of Freedom*:

> …the relatively specific maxims on which agents usually act, for example, the maxim of false promising, presuppose more general principles in the sense that a commitment to the maxim entails a commitment to the more general principle. (Henry Allison, *Kant's Theory of Freedom*. *op. cit.*, p. 94.)

[xlviii] Wood has a good discussion of this issue in *Kant's Ethical Thought* (in which he appeals to Hilly Flitcraft; see Allen Wood, *Kant's Ethical Thought*, *op. cit.*, chapter 3 section 7.3).

[xlix] This, of course, is intended to echo one version of Hegel's famous "empty formalism" objection to Kant's ethics. For a helpful examination of this objection, see Allen Wood, "The Emptiness of the Moral Will," *The Monist* vol. 72, No. 3 (July, 1989), pp. 454-483.

[l] Korsgaard, who (as noted above) defends the claim that positive duties can be derived from the universalizability tests, argues that "in the sort of theory I am defending, responsibility in general is going to look a lot more like responsibility for omission" (Christine Korsgaard, *Self-Constitution: Agency, Identity, and Integrity*, *op. cit.*, p. 175). If Korsgaard is correct about this (i.e., if she is correct that on her theory, responsibility in general tracks positive rather than negative duties) and if the arguments in this paper work, then it looks like her theory might face a problem. How serious this problem is will depend on whether she has resources other than the universalizability tests that can be used successfully to derive positive duties.