Cooperation – Kantian-Style

Jan Willem Wieland
j.j.w.wieland@vu.nl

Abstract

Should you reduce your energy consumption? Tragically enough, it may be better for you, and for everyone involved, to refrain from doing so even if you care about the climate. Given this tragedy, why cooperate? This paper defends the view that not cooperating is morally problematic because it is not universalizable (in a Kantian sense). That is, I will argue that we have universalizability-based reasons to cooperate as long as we have a preference for "collective success" (e.g. a sustainable planet). The problem is that defectors let others fix the problem for them, and in this way make an unfair exception of themselves. Moreover, even when selected agents might not share this preference, they still have to cooperate for the sake of others.

Keywords: climate change; tragedy of the commons; cooperation; universalizability

Inquiry, forthcoming

1. Tragedy of the commons

Is it permitted to drive a gas-guzzler? Either too many others drive such a car, or enough others take more sustainable options. If too many others drive a gas-guzzler, and our city will be polluted anyway, it is better for me to drive a polluting car too. If in such a situation I were to travel more sustainably, I would only be wasting money and time. If enough others travel sustainably, and our city will be clean anyway, it is still better for me to drive a polluting car. No one will be worse off, and I can avoid costs. Either way, it is better for me to drive a polluting car.[1]

	Too many others drive	**Enough others do not**
I drive a polluting car	Polluted city	Clean city
I travel sustainably	Polluted city + cooperation costs	Clean city + cooperation costs

The problem generalizes. Thus Parfit: "There are countless other cases. It can be better for each if he adds to pollution, uses more energy, jumps queues, and breaks agreements; but, if all do these things, that can be worse for each than if none do." (1984: 61-2) To explain the dilemma in general terms, let us use the terms "cooperating" and "defecting," where defecting is consuming more than your fair share of the carbon budget (for example), and cooperating is not consuming more than your fair share.

	Too many others defect	**Enough others cooperate**
I defect	3 Collective failure	1 Collective success
I cooperate	4 Collective failure + cooperation costs	2 Collective success + cooperation costs

[1] Even though pollution comes in degrees, and a single ride can add to it, it arguably will not make anyone worse off. For this presentation of the tragedy, cf. Parfit (1984: 56ff) and the subsequent debate. Following this debate, I will focus on individual citizens, though much will carry over to corporations and governments.

Importantly: I strongly prefer collective success over collective failure (e.g. a clean city over a polluted one). The cooperation costs are far smaller than the costs involved with collective failure. In terms of the matrix, I prefer outcomes 1 and 2 over outcomes 3 and 4, and the difference between 1 and 2 on the one hand, and 3 and 4 on the other is substantial.[2] Even so, I cannot realize collective success by simply paying cooperation costs. Collective success depends on whether *enough* agents pay such costs, not on whether *I* do so. Moreover, I do not want to pay unnecessary costs, and hence prefer 1 over 2, and 3 over 4. That is, regardless of what others will do, it is better for me to defect.

The tragedy is that the same reasoning holds for *all* other agents involved, and, if they act accordingly, we end up with collective failure. The question is: given that it is always better for any individual to defect, why cooperate? Why pay the cooperation costs if doing so "seems like a mere waste of [my] efforts" (Nefsky 2015: 257)?

In the literature, we find at least two prominent types of solutions. First, participation theorists say that we should not participate in the group of defectors (e.g. Parfit 1984: ch. 3, Kutz 2000). Second, consequentialists may point out that this presentation of the decision situation is incomplete, and that there may still be a chance—however small—that cooperating will make a difference to collective success (cf. Kagan 2011, Pinkert 2013). Such solutions are promising, though also controversial (cf. Nefsky 2019), and in this paper I will pursue a different and underexplored approach: Kantian universalizability. The basic idea will be that:

> *Defecting is morally problematic because you want others to cooperate and you should not make an unfair exception of yourself.*

In essence, I think this simple answer is deeply intuitive and understandable to everyone. This paper will defend it (with certain qualifications).

The plan is as follows. I start by presenting my Kantian account (§§2-3), which utilizes Christine Korsgaard's interpretation of Kant's formula of universal law (FUL). I will add to Korsgaard in four main respects: I apply the account to the tragedy of the commons (§2); I defuse an objection by Julia Nefsky (§3); I contrast my account to an alternative Kantian account by Maike Albertzart (§4); and finally I respond to two objections concerning hypothetical reasoning and empirical input (§§5-6).

[2] For example, if we assign a value of −1 to the cooperation costs and −1000 to collective failure, then we get the following ordering: 1 (0) > 2 (−1) > 3 (−1000) > 4 (−1001).

2. Practical contradictions

I will suggest that defecting is problematic much in the same way that certain Kantians consider lying to be problematic. In particular, here is Korsgaard's (1985) analysis of Kant's classic case:

1. *Maxim*: "To get money, I will make a false promise that I will pay my debts later."
2. *Universalization*: Imagine a world where everyone lies to get money.
3. *Test*: In this hypothetical world, can I still achieve my goal (get money) by taking the given means (lying)? No, in a world where everyone lies, no one would believe me, and I would not be able to get any money.

Korsgaard's interpretation of FUL is very influential, and applies to numerous cases. For example, stealing is morally problematic because, in a hypothetical world where everyone steals, others steal from you too, and you cannot make a living and avoid working. In abstract terms, your maxim "to achieve goal G, I will do action A" faces a practical contradiction when you can no longer achieve G, by doing A, in a hypothetical world wherein everyone takes A to achieve G.[3] The underlying diagnosis is this: if you run into a practical contradiction, you are unfair. You want others to act differently (e.g. not to steal from you), but make an exception of yourself for no good reason. As Korsgaard puts it:

> "What the test shows to be forbidden are just those actions whose efficacy in achieving their purposes depends upon their being exceptional. … Intuitively speaking, the test reveals unfairness, deception, and cheating." (1985: 36)

Korsgaard does not address the tragedy of the commons, but her account can be employed to offer a straightforward analysis of it. Generally, if you are free riding on others by not doing your part, then you are assuming that enough others will *not* free ride, and so—I would add—you are implicitly assuming that you are more important than others. That's what is morally problematic.

[3] This test is called in full: "the practical contradiction interpretation of the contradiction in conception."

Note that I will not claim that this is how we should interpret Kant.[4] Instead, my claim is that Korsgaard's account is attractive from a systematic perspective, since it offers an elegant explanation for why one should cooperate. Let me generalize Korsgaard's account as follows:

1. *Maxim*: "To gain some personal benefit, I will not cooperate."
2. *Universalization*: Imagine a world where everyone defects (or more precisely: where defecting is the standard means to gain that personal benefit).
3. *Test*: In such a world, I cannot free ride on the cooperation of others, and gain my benefit (practical contradiction).

What goes for this general maxim, also goes for particular instances such as: "to enjoy my Sunday afternoon, I will drive my gas-guzzler." For, imagine a world where everyone drives—and always has driven—gas-guzzlers. In such a polluted world, I can no longer enjoy my Sunday afternoon.

Again, there is a practical contradiction because I can no longer achieve my goal (enjoy myself) in a world where everyone acts like me (gas-guzzles). Importantly, for this account it is not relevant whether the cooperation of other people is beneficial to me. It is not simply that I prefer others to cooperate (although I prefer that too). It is rather that, by defecting, I allow myself something that I would not allow others to do. Defecting is problematic because I would be making an unfair exception of myself, and assume that others will fix the given problems (here: pollution and climate harms) for me.

Gunnemyr worries:

> "However, my aims would not be frustrated in a world where no one refrains from reducing their emissions of greenhouse gases… Most climate-change-related harms, at least those severe enough to threaten to frustrate any important aim I have, are likely to occur long after my death." (2021: 49)

[4] Kant e.g. wrote: "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" (*Groundwork* 4:424). Sensen (2023) emphasizes the significance of this passage for understanding FUL. Kant's writings on FUL have led to widespread interpretations, each with its own merits (cf. Galvin 2009). The paper's main ambition is to revisit the merits of Korsgaard's account in the context of the tragedy of the commons.

In response to this, it is important to make some assumptions about the timescale of the universalized world. To check if gas-guzzling is (truly) universalizable, we imagine that many past generations drove polluting cars as well.[5] Moreover, even if only future generations suffer from widespread defection in the actual world (cf. Gardiner 2002), I do suffer from it in such a hypothetical—truly universalized—world too. Note that if the practice of gas-guzzling, on its own, were not sufficient for pollution or climate harms, then we may also test the maxim "to have a nice life, I will adopt a carbon-intensive lifestyle." A world where previous generations adopted such a lifestyle would be unliveable, and in such a world I cannot have a nice life (practical contradiction).

Problem solved? This solution to the tragedy of the commons seems almost too simple. At any rate, universalizability-based approaches like this have been met with serious criticism. Sinnott-Armstrong writes:[6]

> "My goals would be reached completely if I went for my drive and had my fun without expelling any greenhouse gases. This leaves no ground for claiming that my driving violates [FUL]." (2005: 303)

It is right that my goal is to have fun and enjoy my Sunday afternoon, and not, for example, to pollute or expel greenhouse gases. However—and this is the crux—that goal *cannot* be reached without collective success. I cannot enjoy my Sunday afternoon when the environment is polluted.[7] I can benefit only if enough others do *not* drive gas-guzzlers.

Next, consider the following challenging comment by the early Parfit:

> "This belief may also undermine the Kantian solution. If my contribution [or cooperation] would make no difference, I can rationally will that everyone else does what I do. … Since others may think like me, it is of great importance whether, in such cases, we can explain why we should contribute…" (1984: 67)

[5] Inspired by Glasgow's idea of temporal universalization (2003).

[6] Tiefensee concurs: "Adopting a Kantian approach, it is not unlikely that we can formulate the maxim on grounds of which Charles acts in such a way that renders his pleasure drives universalizable." (2019: 236) As does Fanciullo: "If the maxim of this person's action was, for example, "don't contribute when doing so would make no difference to the relevant outcome," there would be no contradiction if everyone were to adopt the same maxim." (2021: 432)

[7] Unless I have non-standard preferences, which I address in §3.

Parfit suggests that my maxim could well be: "to avoid wasting costs, I will not cooperate when doing so makes no difference." In a world where everyone avoids actions that make no difference, I can still avoid wasting costs, and there seems to be no practical contradiction. Kutz voices similar scepticism about FUL, and illustrates it based on the following maxim:

> "I will drop my incendiary bombs on the city, in order to avoid the criticisms of my commander and fellow crew, but only because I know these few bombs won't make a difference to whether a firestorm arises." (2000: 134)

Here, too, the agent does not seem to run into a practical contradiction, for it is still possible for her to avoid criticism in a world where others also simply follow orders. But, if this is so, the test fails to explain why it is problematic to defect (here: drop one's bombs). What goes wrong in these cases? This paper's ambition is to respond to this problem.

One important caveat: we will not address the issue of maxim formulation in general, and discuss *all* alleged false positives and negatives of Korsgaard's account. One promising strategy is to reformulate so-called puzzle maxims, i.e. in a systematic way. False negatives like "to avoid crowded courts, I will go to the tennis club on Sunday morning" (cf. Herman 1993: 138) may be restated, for example, as "I will coordinate my hobbies with others, e.g. play tennis when others prefer not to play." In a world where everyone coordinates their hobbies, I can still avoid crowded courts. False positives like "to concentrate on my philosophy paper, I will kill people that make too much noise" (cf. Korsgaard 1985: 28) can, for example, be rendered as "I will use others as a mere means, i.e. without their consent." In a world where everyone does this, I will be used as a mere means myself, and no longer be able to concentrate on my philosophy paper. Such reformulations, for sure, raise further questions that we cannot address here (for the debate, cf. Galvin's 2009 overview). In this paper, we will restrict ourselves to tragedy of the commons cases (i.e. a subset of false positives).

3. Maxim formulation

Let us consider the following case (edited from Parfit 1984: 76):

> **Drops of Water**

> 1000 wounded men are lying out in the desert, suffering from intense thirst. An equal number of people have a pint of water. They could pour these pints into a watercart, which will then be driven to the desert, and the water will be shared equally among the wounded men. Each person who donates a pint enables each wounded man to drink only slightly more water, even less than a drop. One such drop will not benefit even a very thirsty man.

In terms of a decision matrix, the dilemma is as follows:

	Too many others keep it	**Enough others donate**
I keep my pint	Wounded men suffer	Wounded men are helped
I donate my pint	Wounded men suffer + no pint for me	Wounded men are helped + no pint for me

Again, we might think that defecting is not universalizable. For, if all defect (keep their pint), we end up with collective failure. Is keeping one's pint indeed non-universalizable?

Note, firstly, that there is a key difference to the tragedy discussed earlier. In that case, the agent herself benefits from collective success (i.e. clean city). In Drops of Water, only a group of *others* benefit from collective success. Inspired by Kant (e.g. *Groundwork* 4:424), we may think that there is an asymmetry between these cases, and that there are only imperfect duties to help others and, for example, share our resources in this case. For the time being, though, we will focus on this case, as Nefsky poses her challenge in terms of it: "it is far from clear how [FUL] would yield the result that you ought to add your pint to the cart." (2015: 267)

Briefly, Nefsky's reasoning is this. According to FUL, you should donate your pint only if a maxim of "keeping my pint"[8] leads to a contradiction in conception (CC) or a contradiction in the will (CW). Firstly, there is no CC, for "in a world in which everyone has the same maxim" it is still possible "to act successfully on the maxim" and to keep your pint. Furthermore, there is no CW because willing the maxim "keeping my pint" as a universal law is *not* in contradiction "with something else that I must will as a rational agent." As Nefsky explains: "There is no such contradiction. *I* will not be negatively affected if no one puts water into this cart." (ibid.)

[8] Strictly speaking, this is only the act description. More complete maxim formulations will be discussed below.

Still, Nefsky admits that we do obtain a Kantian contradiction if we describe the maxim in a more general way, namely as "refraining from helping others in need." Indeed, this is Kant's own example of a maxim—the maxim of non-beneficence—that yields a CW. For, willing this maxim as a universal law *is* in contradiction with something else that I will (or must will as a rational agent), namely that *I* will be helped when I am in need myself. However, in a world where no one helps others, I will also not be helped when I am in need, and we get the contradiction. At this point, however, Nefsky objects:

> "But here—just when it seems we have found the contradiction we were looking for—we hit the Superfluity Problem. We can only describe my maxim of refraining from adding my pint as a maxim of "refraining from helping those in need" if my adding the pint would help those in need. But the claim that my act won't make any difference to those in need seems to be precisely a claim that it will not help them." (ibid.)

Donating one's pint does not amount to helping the wounded men. This much is given by the case description. But, how does it follow from this that we cannot describe the maxim as "refraining from helping those in need"? The thought seems to be this: donating does not amount to helping, and so, similarly, *not* donating does not amount to *not* helping. Furthermore, if not donating does not amount to not helping, then the description "refraining from helping those in need" is inapplicable.

Yet, this objection is not successful. If keeping one's pint fails to make a difference, it also fails to help the wounded men, and "refraining from helping those in need" can still be a relevant maxim to test.

In broad outline, the Kantian described by Nefsky reasons as follows: I should donate my pint because refraining from helping people in need yields a contradiction. Contrast this to the following reasoning: I should donate my pint because I should help the wounded men, *and donating my pint will actually help them.* The latter would indeed assume that my action will be helpful (and go against the case description without any justification[9]). However, the former reasoning does not make use of such an assumption. Just like we put "making a false promise" to the test, we can test "refraining from helping

[9] Nefsky (2017) herself denies that difference making is required for helping. On her view, donating one's pint can still be described as "helping the wounded men" (i.e. if certain conditions are met). It is important to see that Nefsky's own view is very different from the view explored here. Most importantly, there exist no helping-based reasons—though there should be universalizability-based reasons—when enough people cooperate and collective success is already guaranteed.

people in need," and if the latter runs into a contradiction, it is simply wrong to act on it. (We can also test "speaking truthfully" or "helping the wounded men," though these will simply pass.) One may think: why should I even regard donating my pint as trying to help them? But this question need not arise for the Kantian under discussion.[10] What matters is that *refraining from* helping the wounded men is not universalizable and morally problematic.

Kantian ethics is primarily interested in the agent's plans, intentions or maxims, rather than actual difference making. The prime question is whether you act from objectionable plans. What are you trying to achieve? Imagine that you want to donate your pint to help the men in the desert, but you accidentally spill it on the ground before you are able to add it to the watercart. That may be clumsy or stupid, but your plans are not objectionable. In contrast, if in Drops of Water the agent does not want to help out, then the maxim of non-beneficence could well be a relevant maxim to test.

Yet, a subsequent issue arises: why should the maxim for example be formulated as the maxim of non-beneficence (i.e. one that fails the test), and not in some other way (i.e. one that passes the test)? Or again: *is there any non-arbitrary way to test maxims in tragedy of the commons contexts?* In the following, therefore, I will conduct a more systematic survey of the possible maxims that might be relevant in Drops of Water. Why do I want to keep my pint rather than add it to the watercart? Here is a list of possible goals I could be trying to achieve:

(1) avoiding wasting my pint;
(2) causing collective failure;
(3) enjoying the pint myself;
(4) making an impression on others;
(5) nothing really, just spending my time.

Let us consider these in turn, and see whether they yield a practical contradiction.

Maxim (1): "to avoid wasting my pint, I will not donate it." This is the maxim of a "sophisticated" agent (imagined by Parfit) who understands the dilemma she is in. That is, this agent knows that the situation of the wounded men depends on what a sufficiently large group of agents does, and not on what she individually does. It is an instance of "to avoid wasting costs on something that makes no difference, I won't cooperate," and does not

[10] Albertzart (2019) does describe such a Kantian, and I discuss her account in §4.

seem to yield a practical contradiction. After all, in a world where everyone defects, it is still possible to avoid wasting cooperation costs.

In response, we should say that this agent is not sophisticated enough. After all, avoiding cooperation costs is not the only thing agents want in a tragedy of the commons: they also want collective success. Yet, in a world where everyone defects, this is no longer achievable. One may worry that this preference for collective success was not included in the description of the maxim. In response to this, I will propose two general instructions for formulating maxims of the form "to achieve G, I will do A" in this context:[11]

> **Instruction 1:** Include in the description of "A" only whether the agent is defecting or cooperating, in the specific sense of the given scenario.
>
> **Instruction 2:** Include in the description of "G" all relevant concerns the agent has while doing A, and, if applicable, rank her preferences.[12]

Instruction 1 is fairly straightforward. In Drops of Water, for instance, we do not describe the cooperating act as "lifting one's arm" or "spending energy for 2 seconds" or even as "doing something altruistic," but as "donating one's pint" (as opposed to "keeping one's pint"). In the universalized world, then, we do not imagine a world where everyone lifts their arms, or spends energy for 2 seconds, or does something altruistic, but a world where everyone donates their pint.

Note also that we describe the act from the agent's perspective: whether she thinks she is cooperating or defecting. Typically defectors would describe their conduct as "keeping my pint." However, if some agent is confused, and truly believes she is donating her pint—when she's not—the maxim is still one of "donating my pint." (Compare our earlier example: if you want to donate your pint, but accidentally spill it on the ground, then your maxim is still one of "donating.")

The key instruction is no. 2. Defectors may pursue various goals. They may want to drink the pint themselves, keep it for later, give it to their friends, use it for their plants,

[11] These instructions are not meant to be exhaustive, yet these will suffice for our analysis of the tragedy of the commons.

[12] This instruction is inspired, in part, by O'Neill's comment that "none of the maxims on which an agent acts is irrelevant to assessing the moral status of his acts" (1975: 109). Sandler (2010: 173-4) argues that Kantians have difficulties with unintended side-effects (like climate harms). This seems to be less of a problem if we also include the implicit goals of the agent, in addition to her explicit ones (cf. Korsgaard 1985: 41-2).

and so on. In the case of sophisticated agents, we said, two goals are relevant: they are first of all interested in securing collective success—their top priority—but also, though secondarily, in avoiding cooperation costs. Hence, according to instruction 2, their maxim may be restated as follows: "to avoid wasting cooperation costs while, above all, keeping collective success on the table, I will not cooperate." This maxim yields a practical contradiction, because, in a world of universal defection, collective success is no longer on the table.

What if someone is *really* convinced that donating her pint is inefficacious (and doesn't benefit anyone), and keeps her pint *only* because of this, and not because she is selfish in any way? We still say to this person: your maxim is not universalizable. If you care about collective success (i.e. helping the men in the desert), which you do, then your concern is frustrated in a world where people act like you (i.e. keep their pints too). Inspired by Nefsky's worry mentioned above, one may wonder: "If the sophisticated agent does not think that either adding the pint or not adding it will help or hurt collective success, then the latter would not be part of what she is trying to achieve in keeping her pint." In response, I agree that the sophisticated agent might not be trying to secure collective success. It merely holds that this central concern of hers will be frustrated, i.e. after universalization.

To be sure, there may be selected agents who do not care about collective success, and we will consider such agents next. But, in those cases it is important to see that there is no tragedy to begin with. There is a tragedy of the commons *only if agents strongly prefer collective success, but end up with collective failure*.

Maxim (2): "to let the wounded men suffer, I will not donate my pint." This is the maxim of an ill-willed agent, and seems to pass the test. For, in a world where no one donates, the wounded men suffer, I actually *achieve* my goal. The same holds for its more general counterpart: "to cause collective failure, I will defect."

Maxim (3): "to enjoy the pint myself, I will not donate it." This is the maxim of an egoistic agent, and, again, seems to pass the test. For, in a world where others keep their pint too, I can still enjoy my own pint. Of course, I may feel bad about the men in the desert, but I may also try to ignore them and focus on other things. In the latter case, there is no practical contradiction. Importantly, this holds only in such "collective benefit" cases. In the "collective harm" case we saw earlier, the maxim "to enjoy my Sunday afternoon, I will drive my gas-guzzler" does run into a practical contradiction, because I cannot enjoy my Sunday afternoon in a polluted world. So, generally, egoistic agents will run into a practical contradiction when they themselves need collective success to gain their personal benefits.

Maxim (4): "to make an impression on my friends, I will not donate my pint." This is the maxim of an agent who is concerned about her reputation. Does this pass the test? In a world where all keep their pint, it still seems possible to make an impression on one's friends.[13] The same goes for kindred concerns, like preserving one's self-image. Perhaps you have never donated anything in your life thus far, and you are not strong enough to face this fact about yourself, and thus you cook up stories—rationalizations—on why not donating anything is morally fine. For example, you may tell yourself that the men in the desert brought it upon themselves, or that you do not need to help because enough others will probably do so. In a world where everyone lives in ignorance, it seems still possible to preserve your self-image.

Maxim (5): "to spend my time and actually for no real reason, I will keep my pint." This is the maxim of an indifferent agent who could not care less about anything. Compare: "to get through my days, I will drive a gas-guzzler." Such maxims do not seem to yield a practical contradiction. In a world where defecting is universal practice, it seems that I can still get through my days. I just could not care less about anything, and then it does not matter if others also keep their resources to themselves, or pollute the environment. I do not make an exception of myself, at least not in the sense that I want others to fix any problems for me.

All in all, in selected cases (i.e. of ill-will, egoism, image, or indifference), we have seen that it is not so clear that defecting leads to a practical contradiction (and that acting on such motives is problematic in that way). What could Kantians say about such cases? There are at least three points to be made.

First, and most importantly: following instruction 2, we may ask: is the goal formulated adequately? In the case of ill will, for example, one may wonder: *why* does she want the wounded men to suffer? Perhaps this agent primarily seeks attention or love. In that case, it is not so clear that the agent will be able to avoid a practical contradiction. For, in a world where no one donates, or pays any attention to others, it is not so clear that the agent will still be able to receive the attention or love she seeks. Also, in the impresser case, one may wonder whether reputation is really the *only* thing that the agent is trying to achieve. Is she not interested in collective success as well? If so, she will run into a practical contradiction after all.

Second, maxims of defection that pass the test may be rare in real life. As Korsgaard (1985: 43) speculates: "[the temptation to make oneself an exception] and not

[13] Though one may wonder whether keeping one's pint still attracts any attention, i.e. when it is universal practice.

[violent crimes motivated by despair or illness] is the sort of evil that most people are tempted by in their everyday lives."

Third, even if selected defectors do pass Korsgaard's test (and really make no exception of themselves in this particular sense), there may still be a *further* Kantian analysis of what is problematic about them (i.e. they may still make an exception of themselves in some other sense).[14] Specifically, as I will discuss in §4, these selected defectors may still have to cooperate for the sake of *others*.

To sum up, the main result of this section is this: agents, as far as they find themselves in a tragedy of the commons, care about collective success, and as long as we include this in the description of the maxim (as per instruction 2), defecting is not universalizable. For, if others were to defect (which they actually may or may not do), then I would no longer get what I want: collective success.[15]

The following question presents itself: can one now defect (following the proposed account), if one simply drops one's concern for collective success? More generally, does this account reduce Kantian reasoning to mere "hypothetical reasoning"? I will respond to this concern in §5.

Let me finally note that some Kantians suggest that we should not conceive of maxims as specific intentions (of the form "to achieve G, I will do A"), but rather as more general policies (of the form "in circumstances C, I will do A").[16] This distinction is important, though it does not really seem to impact my solution. After all, if specific intentions (like "to enjoy my Sunday afternoon, I will drive a gas-guzzler") run into a practical contradiction already, then more general policies of defecting in multiple contexts (such as "whenever I am in a tragedy of the commons context, I will defect") are even more likely to fail the test.

---

[14] Korsgaard, too, employs two different tests. As she sees it (1985: 39-40), in a CC the agent can no longer achieve the goal specified in the maxim, while in a CW the agent may still be able to achieve that specific goal, but is no longer effective in this (or free in setting her goals in the first place). In a world where no one helps others, then, it may be possible to take my ease, but I will not be effective (i.e. if no one assists me).

[15] This may also apply to Kutz's firebomber (cited in §2): in a world where everyone uncritically accepts the orders from their commanders, avoiding the firestorm is off the table. What if I know that too many others won't go against their commanders? See §6 below.

[16] Cf. O'Neill (1975), Braham & Van Hees (2015), Nyholm (2017), among others. Note that problems arise if we would admit maxims with very restricted circumstances, such as "I will defect, but only when I turn 100 years old"—though see McCarty (2015).

4. Alternative account

In the following, I will further clarify my account by contrasting it to a recent alternative account by Albertzart (2019). Basically, Albertzart's proposal is that we can derive an imperfect duty to cooperate (e.g. to refrain from gas-guzzling) from the imperfect duty to help others. This proposal is worth considering as it offers a second response to Nefsky's challenge (and will supplement my account, as will become clear). The question is why cooperating (e.g. adopting a less carbon-intensive lifestyle) would count as "helping others" if doing so makes no difference. Albertzart aims to account for just this.

According to Kant, you have an imperfect duty to help and contribute to the well-being of others (*Groundwork* 4:423-4, 430). This implies that you have some latitude regarding what to do exactly, and when to do it. The same does not hold, for example, for the perfect duty not to keep slaves for personal gain. You simply should never do this. Importantly, an imperfect duty to help does not imply that it suffices to, say, donate a pint just once in your life. The imperfect duty to help others can still be very demanding.[17] In particular, Albertzart argues that this duty is incompatible with a carbon-intensive lifestyle. Genuinely embracing the end of helping others does not come cheap.

How does Albertzart get this result? In total, there are a few steps: from the imperfect duty to help others, to selecting the necessary means to help others, to combating climate change, to selecting certain (sufficient) means to combating climate change, to avoiding unnecessary car trips.

Let us briefly go through these steps. The first step: "To adopt the happiness of others as an end implies willing the necessary means for achieving this end. Given the negative impact climate change is expected to have on human happiness, combating climate change qualifies as one of these necessary means." (2019: 844) This step is fairly straightforward: if we should help others, and combating climate change is necessary for this, then we should want to combat climate change (i.e. together). The next step: "combating climate change … in turn, can be secured through different lower-level means … Because there is ... more than one effective means to combating climate change, the lower-level means an agent has to will depend on the higher-level means she chooses." (ibid.) This step is slightly more complex: combating climate change requires one to select

[17] The exact details form a central issue in Kantian ethics, cf. e.g. Stohr (2011).

certain means to achieve this (again, together).[18] For example, one strategy would be to choose the means "everybody avoids unnecessary car trips." Based on this example, Albertzart explains how such trips run into a Kantian contradiction:

> "Assume the … motorist wills that everybody avoids unnecessary car trips as a means to combating climate change. In this case, she cannot allow herself such trips. She faces a contradiction between a chosen means for an obligatory end—namely, that everybody avoids unnecessary car trips—and the universalised form of her maxim, namely, that everyone drives their cars in order to reach their destination conveniently, efficiently, and in privacy." (ibid.)

As Albertzart sees it, choosing certain means to combat climate change does not necessarily mean that you should actually choose the means described here ("everybody avoiding unnecessary car trips"). You may also choose everybody becoming vegan, or everybody boycotting flights, or everybody protesting government policies and demanding institutional changes. As Albertzart clarifies, "it is up to each individual agent herself to choose the lower-level means," and "agents do not need to discuss and agree upon the specific lower-level means" (ibid.).

This latter aspect of Albertzart's approach invites a worry. If you should choose certain means to combat climate change, then which means will do? Should these in fact be sufficient for combating climate change, or merely be considered sufficient by the agent? If the former, then "everybody avoiding unnecessary car trips" might not do (without additional measures like institutional changes). If the latter, then even "everybody avoiding unnecessary emails" might suffice. Albertzart seems to suggest the latter when she says that the lower-level means are "up to each individual." However, such means would not generate the above contradiction, and entail that driving gas-guzzlers is wrong. To avoid this implication, I think we may want to restrict the range of lower-level means that people are permitted to select.

Regardless of this detail, it is instructive to contrast Albertzart's approach to mine. Following Albertzart's account, we should cooperate because we should help others, and we should not let others fix this for us, and be unfair in this way. Following my proposal, we should cooperate because we should not let others fix our *own* problems, and be unfair

[18] Refusing to choose any such means, Albertzart adds (ibid.), amounts to a violation of a perfect duty. That is, the imperfect duty to help others entails a perfect duty to select certain sufficient means to combat climate change.

in this alternative way. Even though both constitute Kantian universalizability-based accounts, they are distinct. Let me point out two key differences. First, Albertzart's account relies on the assumption that we should help others (specifically, to adopt "the obligatory end of the happiness of others"). You run into a contradiction, on her proposal, if you leave the work of meeting this duty to others. My account makes no assumption about whether we should help others. It merely uses the agent's own maxim (formulated according to the two instructions) to derive a contradiction. You run into a contradiction when you frustrate your *own* concerns.

Second, Albertzart defends an imperfect duty to cooperate (with some latitude for agents to comply with it), while I defend a perfect duty to cooperate.[19] This difference may appear merely terminological. After all, one may think, even when you have a perfect duty not to take more than your fair share (e.g. of the carbon budget), you still have some latitude in taking your share as long as it is not *more* than your fair share.

But, the difference doesn't seem merely terminological. As Albertzart (2019: 840) points out, driving gas-guzzlers is only extrinsically problematic. Doing so is wrong only in a context where others drive gas-guzzlers and where you together cause climate harms.[20] Inspired by Kant, to some extent I want my account to be insensitive to such empirical input (though I am going to qualify this in §6). Just as lying for personal gain is problematic even when no one in fact lies, I think that defecting in a tragedy of the commons is problematic even when no one in fact defects. That is, even when collective success is not at risk, defecting can still be morally problematic, i.e. for universalizability-based reasons.

One may point out that one's "fair share" will itself be defined, at least partly, based on how people actually behave. Consider one of Albertzart's cases: picking wild mushrooms (2019: 841). Regulations to preserve the mushroom population differ regionally. One regulation is that picking wild mushrooms is prohibited between the first and the tenth of each month, and that you cannot take more than ten kilos on other days. Another regulation is that you are not allowed to pick more than 250 grams each time, regardless of the day of the month. Hence, there are several distinct ways to determine one's fair share of, in this case, wild mushrooms.[21] Furthermore, such regulations already

[19] I argued that maxims of defection can fail the CC-test, and we have perfect duties not to make an exception of ourselves in this sense.

[20] According to Albertzart (2019: 840-3), Parfit's (2011) universalizability-based principles might render such extrinsic acts either *all* permissible or *all* impermissible.

[21] We may let this depend on given existing regulations (if they exist) or else on certain ideal regulations. Compare the debate on the "carbon budget" (cf. Baatz 2014). Determining the amount and form of such a budget is, to be sure, complicated, and forms a separate discussion. Davidson

incorporate empirical information about how many people actually want to pick wild mushrooms. The rule of a maximum of 250 grams may work in some regions where only 2% of the population is interested in wild mushrooms, but not if 100% is interested. Furthermore, if your fair share depends on such information, then taking more than your fair share is also only extrinsically wrong.

As I see it, such an approach to determining one's fair share may well be taken on board. However, once we settle on some such measure of fair share, my account says that you have a perfect duty not to take more than it—that is, *even when all others adhere to the regulations and do not take more.*

Note that I do wish to distinguish "collective harm" and "collective benefit" cases. In the latter, a group of agents does not cause harm, but rather fail to benefit others. Even when we have a perfect duty not to defect in collective harm cases (like climate harms), there may still be only an imperfect duty to cooperate in collective benefit cases (e.g. to donate one's pint in Drops of Water). Such an asymmetry is also Kantian in spirit, but how to account for it? In both cases, one might think, I am committed to saying the same thing. In a world of universal defection, your preference for collective success is frustrated. In a world where everyone pollutes, your preference for a clean city is frustrated. Similarly, then, in a world where all keep their pint, your preference for relieving the thirst of the men in the desert is frustrated.

Even so, there is an important difference. In many collective harm cases, your preference for collective success is, to some extent, non-optional. For instance, without collective success you yourself cannot benefit from a clean city. In typical collective benefit cases (such as Drops of Water), you yourself do not benefit from collective success (only a group of others), and at least in this sense your preference for collective success is more optional. As long as you do not care about the wounded men in the desert, you do not run into a practical contradiction if you keep your pint.[22]

My account does not explain why you should have certain concerns and preferences in the first place.[23] After all, you may also be completely indifferent, and not

(2023) suggests that we should base our decisions on a "shadow price" for carbon: "if we would not go out for a drive on a sunny Sunday afternoon in a gas-guzzling sport utility vehicle if gas prices were twice as high, we should not do it now."

[22] Strictly speaking, the difference is between harming or benefiting *others* as opposed to harming or benefiting *yourself*.

[23] In the context of a different approach, Soon, too, focuses exclusively on, what she calls, the "Ordinary Person": "Unlike someone who does not care at all about collective harm, the Ordinary Person is committed to the good." (2021: 3355)

care in the slightest about polluted cities—that is, even when you have to live in them yourself. It is questionable that there are many such people.[24] In any case, we need an additional story such as Albertzart's to address these concerns. My account explains why we should cooperate *given certain preferences we have*. In addition to this, Albertzart offers an account of why we should care about collective success in the first place: if not for ourselves, we should care about it for others.[25]

For sure, the details of the overall account (based on these two universalizability-based arguments), as well as the exact nature of the resulting duties, merit further attention. Perhaps we would need a more sophisticated framework, i.e. that goes beyond the perfect/imperfect distinction (cf. Lichtenberg 2010, Sticker 2023). I will use the remainder of the paper to address two important concerns about my account.

5. Hypothetical reasoning

The first concern is as follows: "You suggest that we should cooperate as long as we have a preference for collective success, but shouldn't our duties to cooperate be *independent* of any contingent preferences we happen to have?"

In response, let me start by pointing out that my position here is in line with Korsgaard's account. On her view (1985: 40-3), while FUL can account for maxims like "to get the job, I will kill my rivals" (if everyone does this, I will be killed too and not get or keep the job), FUL does seem to have its limits. Specifically, Korsgaard points out, it has a harder time with maxims like "to satisfy my hatred, I will kill someone." This maxim avoids a practical contradiction if the wrongdoer is ill and primarily cares about the immediate result (here: someone's death). The same applies in the tragedy of the commons: if the agent is truly indifferent to collective success, then the agent does not assume that others will fix it for her and make an exception of herself in this way.

Still, you may wonder: doesn't this reduce everything to hypothetical reasoning? Indeed, we are not interested in hypothetical but *categorical* output of the test (cf.

[24] In fact, Kant himself might have thought that most people want this. For, Kant held that happiness is a shared end that people have in virtue of their human nature (e.g. *Groundwork* 4:415, cf. Nyholm 2015, Bojanowski 2017). If collective success is required for happiness, then people—as far as they are rational—will also have shared desires for collective success.

[25] Albertzart only addresses collective harm cases, though her account naturally extends to collective benefit cases. That is, if we should combat climate change for the sake of other people, we plausibly also have to do our part in projects of beneficence.

Groundwork 4: 414ff). Hypothetical imperatives specify necessary means to your ends (cf. Hill 1973). An example is: "if you want to compete for the job, and working hard is necessary for this, then you have to work hard." In the case of such imperatives, you have two options: hold onto your goal and take the necessary means to that goal, or drop the goal in order to avoid the necessary means. You can either compete for the job and work hard, or take a less competitive job.

The question is: how does Korsgaard's account of FUL differ from this? After all, both concern certain means/end combinations. In the case of hypothetical imperatives, you weigh a goal (the job) against the benefits of not having to take the necessary means (work hard), and ask yourself whether the former indeed outweighs the latter. FUL, in contrast, does not weigh ends and means. Instead, it says that, given your ends, you should not take certain means (necessary or other). Specifically, you should not take means when you cannot allow others to take them, on pain of frustrating your goal (and making an unfair exception of yourself in this way). You should not kill your rivals when you cannot allow others to kill *their* rivals (on pain of frustrating your goal to get the job).

If you do not want to take the necessary means, you can drop your goal. In the case of FUL, you may suggest that a similar strategy is available, namely, *to drop the goal when the means are not universalizable* (rather than undesirable). If you really want to kill your rivals, then why not say you are no longer interested in the job?

It is crucial to see that it is not possible to game FUL in this way. First of all, you cannot just *say* you are no longer interested in the job (if you actually are). If you are interested in the job, it should be included in the description of the maxim (as per instruction 2, discussed in §3). If, in contrast, you are more interested in other things (e.g. satisfying your hatred), then *that* should be included in the description of the maxim. Finally, and most importantly, wanting to take certain means that are not universalizable is inherently problematic.[26] That's the whole point of FUL.

All this carries over to the tragedy of the commons. According to my account, it is not that you should cooperate because this would be a necessary means to certain goals of yours, or that you can also decide to drop the goal rather than take the means. Instead, you should cooperate, given certain interests of yours (i.e. collective success), because not doing so is not universalizable. As discussed in the previous section, there may still be rare cases where someone is not interested in collective success. But these are not agents that try to game FUL (i.e. drop the goal when the means are not universalizable).

[26] Moreover, gaming FUL may not be permitted by FUL itself, cf. Sneddon (2011).

6. Empirical input

There is a second concern we should address. Consider the following case from Gruzalski (1982: 33):

> **Dictatorship**
> A million citizens are needed to revolt against their dictator and overthrow the regime. But, acts of revolting are punished with the death sentence by the regime and participants of the revolt are easily tracked down by the secret police. Given the situation, it is certain that most citizens will not speak up, and that there will not be enough people to make the revolt a success. This has very bad consequences for the whole country, including economic breakdown and further loss of liberties.

The question is: does it follow from the account defended that one should revolt, and give up one's life? Note, firstly, that the cooperation costs (death) are higher than the costs of collective failure (no liberties). This differs from the tragedy of the commons cases we discussed, and changes the decision situation. If the cooperation costs are too high, then you might prefer not paying them over collective success, and then you will not run into a practical contradiction (at least you will not frustrate your top priority of staying alive) if you defect.

Therefore, let us imagine a variant of the case where you will not be punished with the death sentence, and that the only cooperation costs involve some time and effort to make a banner and join a demonstration. In that case, not going seems to run into a practical contradiction: if all were to stay home—and not signal to others that they are willing to unite (in terms of Lawford-Smith 2015 and Hindriks 2023)—then the success of the revolt would be off the table. But now what if I know for a fact that not enough people will join the demonstration? Does it follow from my account that, in such a case, I should pay cooperation costs *for nothing*, and waste my time and energy?

Following Gruzalski, some will be tempted to take this as a reductio of universalizability-based approaches.[27] In response, I see several ways to go. The hard-line

[27] For such concerns, cf. Glover (1975: 177), Johnson (2003), Budolfson (2012), Cripps (2013: 136-7), Barry & Øverland (2016: 230), among others. It is important to keep the current worry separate from coordination problems. For example, imagine that only 10% of the population is needed to cooperate and overthrow the regime. In that case, the citizens should coordinate their actions, and these type of cases require a *further* analysis (as noted in §2).

response would be to say: yes, there are universalizability-based reasons to cooperate even when collective success is off the table. The soft-line response would be to admit that there are no such reasons.[28] In the following, I do not wish to offer decisive support for the latter response, though I will suggest that, surprisingly enough, my account can be used to justify it.

Classically, FUL is taken to be insensitive to empirical input about people's actual conduct. Lying for personal gain is impermissible even when, as a matter of fact, no one lies in the actual world. After all, in the test we are merely considering a hypothetical world full of liars, not an actual one. Similarly, then, it should not matter, one may think, if and how many people in fact cooperate or defect (in a certain tragedy of the commons situation). One consideration in favour of this hard-line response has to do with self-legislation (cf. Davidson 2023). Kantians value the idea that you impose your own laws on yourself, and are not dependent on external factors, such as social or legal expectations or natural impulses (e.g. to copy other people's behaviour). Similarly, then, you may think that it is valuable if you let your decision to cooperate *not* depend on other people's behaviour.

As I see it, it is important to accept that morality can demand us to act in certain ways even when no one in fact complies. For example, even in a world where everyone lies to one another, we can still have universalizability-based reasons to refrain from doing so. Hence, it may well be that we have reason to cooperate even when most others do not. Morality may even demand us to cooperate if we are the only one. Even so, this is not the same as saying that you cannot let your decision to cooperate depend on certain things, or that doing so would necessarily be invaluable. What matters, instead, is that you do not make an unfair exception of yourself. Do you commit this mistake if you cooperate only conditionally?

We said that defecting is not universalizable because in a world where everyone defects your preference for collective success will be frustrated. But in the dictator case you know that, in actuality, too many others are already going to defect, and collective success is no longer within sight. In such a situation, you can deprioritize your preference for collective success (as it is not achievable), and prioritize your preference for avoiding cooperation costs. If you do this, moreover, *you no longer run into a practical contradiction if you defect*. If you defect and collective success is still on the table, we said, you assume

[28] This latter response is inspired by the social contract tradition (cf. Gauthier 1987). Verbeek (2002: 106-18), too, suggests that Kantian ethics may be compatible with both conditional and unconditional cooperation (i.e. conditional on the cooperation of others).

that enough others do *not* defect, and in that sense you make an exception of yourself. But this does not apply here. When collective success is no longer within sight, you need not assume that others will fix it for you, nor consider yourself more important than them in this way.[29] In such a case, then, defecting would be permitted, yielding the soft-line response that you should not waste cooperation costs.[30]

	Too many others defect 100%
I defect	Collective failure
I cooperate	Collective failure + cooperation costs

Does the same go for the opposite situation wherein you know that, actually, enough others are already going to cooperate? In a sense, your contribution is similarly wasteful. As Parfit puts it: "I may believe that the case is like those where some threshold has been clearly passed, so that any further altruistic act is a sheer waste of effort." (1984: 67) To illustrate, consider a second case by Gruzalski (1982: 30):

> **Water-shortage**
>
> I live in a small town which must suffer a severe water shortage for ten days. Meanwhile, in order to conserve water so that critical needs are met, each person is required, among other things, not to bathe more than twice in the ten-day period. After six days it is clear that if some, but not all, of us took extra baths, there would be absolutely no danger of producing a shortage for critical needs. Suppose that I can bathe more than twice without anyone in town knowing and that the consequence of doing so will be an increase of my own happiness as well as the happiness of those near to me. Finally, I know my fellow citizens, and know that they will not take any extra baths.

[29] In this case, you deprioritize your preference for collective success because it is no longer achievable, and not because you want to game FUL (on the latter, see §5).

[30] Hindriks (2023) also suggests that cooperation is required only if the "prospect of success is good enough." Note that there may still be *other* reasons to cooperate when collective success is no longer possible. For example, Wieland (2022) argues that, in such contexts, there may still be strong participation-based reasons not to defect (namely, to avoid membership of the group responsible for collective failure).

Hence, the question is: when enough others are already cooperating (here: do not take extra baths), why should you pay the cooperation costs (your smell)? Why not follow Gruzalski's act utilitarian advice of taking the extra bath? In such cases, however, collective success is still achievable. Therefore, you will still run into a practical contradiction if you defect. For, in a hypothetical world where all defect (which is then quite unlike the actual world), your preference for collective success, which you still have, will be frustrated. Thus, when enough others cooperate, you still have universalizability-based reasons to pay the cooperation costs (and to go against the act utilitarian's advice).

This is only a sketch of how the account may be developed further.[31] Let me mention one further point. A familiar concern about Kantian ethics is that it does not care about actual outcomes, but rather about intentions and whether people's minds are of the right sort. In this case, too, one may wonder whether the view is serious enough about collective harms *out there in the world* (cf. Jamieson 2007: 161, Cripps 2013: 137-8). Are universalizability-based reasons the right sort of reasons in this context? This is too big a question to address here, but let me just offer two brief considerations in response. Firstly, the proposal is that you should avoid making an unfair exception of yourself, that is, regarding the solution for exactly those collective harms out there. Secondly, when collective success is not on the table, we just said that there may be no reason to cooperate only to preserve the right mind set. In all other cases, acting on universalizability-based reasons will not only prevent collective failure in hypothetical worlds (where others act like you), but also—if enough people decide to act on them—in the actual world (cf. Goodin & Barry 2021: 346).

7. Conclusion

Let me wrap up. Why cooperate in cases where it is, tragically enough, better for each agent to defect? In this paper, I defended the view that defecting is morally problematic because it is not universalizable in a Kantian sense. This paper's main ambition was to respond to influential worries by Nefsky and Sinnott-Armstrong (among others), and show that the

[31] In this paper, I assumed that universalizability-based reasons are on-off: either you have them, or you do not have them. Yet perhaps it makes sense to think of universalizability-based reasons in terms of degrees, i.e. that one can make an unfair exception of oneself to a greater and lesser degree. One potentially relevant factor may be the strength of your preferences. In Water-shortage, your universalizability-based reason against taking an extra bath may then depend on the strength of your preference for collective success (i.e. water for you and others in the future).

main idea is not only intuitive, but can be defended along the lines of Korsgaard's account. Namely: in a world where everyone defects, you cannot free ride on the cooperation of others, and gain your personal benefit. Or, in more ordinary terms: defecting is problematic because you would be making an unfair exception of yourself and letting others fix the given problems for you (or, following Albertzart, for others).[32]

References

Albertzart, M. 2019. A Kantian Solution to the Problem of Imperceptible Differences. *European Journal of Philosophy* 27: 837-51.

Baatz, C. 2014. Climate Change and Individual Duties to Reduce GHG Emissions. *Ethics, Policy & Environment* 17: 1-19.

Barry, C. & G. Øverland 2016. *Responding to Global Poverty. Harm, Responsibility, and Agency*. Cambridge: CUP.

Bojanowski, J. 2017. Thinking about Cases: Applying Kant's Universal Law Formula. *European Journal of Philosophy* 26: 1253-68.

Braham, M. & M. Van Hees 2015. The Formula of Universal Law: A Reconstruction. *Erkenntnis* 80: 243-60.

Budolfson, M. B. 2012. *Collective Action, Climate Change, and the Ethical Significance of Futility*. PhD dissertation, Princeton University.

Cripps, E. 2013. *Climate Change and the Moral Agent Individual Duties in an Interdependent World*. Oxford: OUP.

Davidson, M. D. 2023. Individual Responsibility to Reduce Greenhouse Gas Emissions from a Kantian Deontological Perspective. *Environmental Values* 32: 683-99.

Fanciullo, J. 2021. The Psychological Basis of Collective Action. *Philosophical Studies* 178: 427-44.

Galvin, R. 2009. The Universal Law Formulas. In *The Blackwell Guide to Kant's Ethics*, pp. 52-82. Malden, MA: Blackwell.

Gardiner, S. 2002. The Real Tragedy of the Commons. *Philosophy & Public Affairs* 30: 388-416.

[32] For comments at various stages of this project, I thank my colleagues and students at Vrije Universiteit Amsterdam, participants of multiple OZSW conferences in the Netherlands (dating back to 2017!), the participants of the "Small Acts, Big Harms" workshop (Helsinki, 2021), two referees of the journal, as well as the editor Martin Sticker.

Gauthier, D. 1987. *Morals by Agreement*. Oxford: OUP.

Glasgow, J. M. 2003. Expanding the Limits of Universalization: Kant's Duties and Kantian Moral Deliberation. *Canadian Journal of Philosophy* 33: 23-47.

Glover, J. 1975. It Makes No Difference Whether or Not I Do It. *Proceedings of the Aristotelian Society* 49: 171-90.

Goodin, R. E. & C. Barry 2021. Responsibility for Structural Injustice: A Third Thought. *Politics, Philosophy & Economics* 20: 339-56.

Gruzalski, B. 1982. The Defeat of Utilitarian Generalization. *Ethics* 93: 22-38.

Gunnemyr, M. 2021. *Reasons, Blame, and Collective Harms*. PhD dissertation, Lund University.

Herman, B. 1993. *The Practice of Moral Judgment*. Cambridge, MA: HUP.

Hill, T. E. 1973. The Hypothetical Imperative. *Philosophical Review* 10: 429-50.

Hindriks, F. 2023. When to Start Saving the Planet? *Journal of Ethics & Social Philosophy* 23: 397-419.

Jamieson, D. 2007. When Utilitarians Should Be Virtue Theorists. *Utilitas* 19: 160-83.

Johnson, B. L. 2003. Ethical Obligations in a Tragedy of the Commons. *Environmental Values* 12: 271-87.

Kagan, S. 2011. Do I Make a Difference? *Philosophy & Public Affairs* 39: 105-41.

Kant, I. 1785. *Groundwork of the Metaphysics of Morals*. Trans. M. Gregor & J. Timmermann 1998. Cambridge: CUP.

Korsgaard, C. M. 1985. Kant's Formula of Universal Law. *Pacific Philosophical Quarterly* 66: 24-47.

Kutz, C. 2000. *Complicity. Ethics and Law for a Collective Age*. Cambridge: CUP.

Lawford-Smith, H. 2015. Unethical Consumption and Obligations to Signal. *Ethics & International Affairs* 29: 315-30.

Lichtenberg, J. 2010. Negative Duties, Positive Duties, and the New Harms. *Ethics* 120: 557-78.

McCarty, R. 2015. False Negatives of the Categorical Imperative. *Mind* 124: 177-200.

Nefsky, J. 2015. Fairness, Participation, and the Real Problem of Collective Harm. *Oxford Studies in Normative Ethics* 5: 245-71.

Nefsky, J. 2017. How You Can Help, Without Making a Difference. *Philosophical Studies* 174: 2743-67.

Nefsky, J. 2019. Collective Harm and the Inefficacy Problem. *Philosophy Compass* 14: 1-17.

Nyholm, S. 2015. Kant's Universal Law Formula Revisited. *Metaphilosophy* 46: 280-99.

Nyholm, S. 2017. Do We Always Act on Maxims? *Kantian Review* 22: 233-55.

O'Neill, O. 1975. *Acting on Principle*. 2nd ed. 2013. Cambridge: CUP.

Parfit, D. 1984. *Reasons and Persons*. Oxford: OUP.

Parfit, D. 2011. *On What Matters*. Vol. 1. Oxford: OUP.

Pinkert, F. C. 2013. *Global Problems and Individual Obligations*. PhD dissertation, University of St Andrews.

Sandler, R. 2010. Ethical Theory and the Problem of Inconsequentialism: Why Environmental Ethicists Should be Virtue-Oriented Ethicists. *Journal of Agricultural & Environmental Ethics* 23: 167-83.

Sensen, O. 2023. Universal Law and Poverty Relief. *Ethical Theory & Moral Practice* 26: 177-90.

Sinnott-Armstrong, W. 2005. It's Not My Fault: Global Warming and Individual Moral Obligations. *Perspectives on Climate Change* 5: 221-53.

Sneddon, A. 2011. A New Kantian Response to Maxim-Fiddling. *Kantian Review* 16: 67-88.

Soon, V. 2021. An Intrapersonal, Intertemporal Solution to an Interpersonal Dilemma. *Philosophical Studies* 178: 3353-70.

Sticker, M. 2023. Poverty, Exploitation, Mere Things and Mere Means. *Ethical Theory & Moral Practice* 26: 191-207.

Stohr, K. 2011. Kantian Beneficence and the Problem of Obligatory Aid. *Journal of Moral Philosophy* 8: 45-67.

Tiefensee, C. 2019. Why Making No Difference Makes No Moral Difference. In *Demokratie & Entscheidung*, pp. 231-44. Wiesbaden: Springer.

Verbeek, B. 2002. *Instrumental Rationality and Moral Philosophy. An Essay on the Virtues of Cooperation*. Dordrecht: Springer.

Wieland, J. W. 2022. Participation and Degrees. *Utilitas* 34: 39-56.