Erasmus Journal for Philosophy and Economics,
Volume 17, Issue 1,
Summer 2024, pp. 126–148.
https://doi.org/10.23941/ejpe.v17i1.786

Intra/Inter Paradox

JAN WILLEM WIELAND
Vrije Universiteit Amsterdam

Abstract: This paper addresses the following paradox. (P1) It is permitted to defect in intrapersonal dilemmas as long as there is a solution to achieve one's long-term goal. (P2) It is not permitted to defect in interpersonal dilemmas, even when there is a similar solution to achieve a collective goal. (P3) There is no relevant difference between intrapersonal and interpersonal dilemmas. At least one of the three propositions must go. In this paper, I show how (P1) is supported by Chrisoula Andreou's work, and offer a defense of (P2). This has surprising implications. The aim will not be to solve the paradox, but rather to show there is one that we should attend to in the first place.

Keywords: intrapersonal, interpersonal, doing in progress, precommitment, fairness

I. THE PARADOX

Compare:

> **INTRA**
> I want to finish my book this year. Any particular day is trivial: if I do not work on it today, I can work on it tomorrow. A problem arises only if I procrastinate too often. There are 365 days, and I know that I may defect—relax and not work on my book—on roughly 40 days. If I defect on too many days, the 'doing in progress' (writing the book) will not be completed. In response to this problem, I make an agreement with my friends that, after 20 free days, I will have to donate €100 to charity each day I give in. I know that this will stop me from procrastinating.

AUTHOR'S NOTE: For comments on this paper, I thank: Tessa Supèr, Chrisoula Andreou, Rutger van Oeveren, Jelle de Boer, Benjamin Mullins, and a reviewer of the journal.

INTER

A single drive in a gas-guzzler has only trivial effects on the environment. If you go for a ride, no one will be harmed more (than they already would have been). A problem arises only if too many people defect and we all do things with too big a carbon footprint.[1] There are 365 thousand of us, and we know that roughly 40k people may defect. If too many people defect, the 'doing in progress' will lead to a destroyed planet. In response to this problem, we decide to make these cars illegal after the production of 20k gas-guzzlers. We know that this will stop us from destroying our planet.

Here's the paradox:

(P1) It is permitted for me to defect, for example on day 15, in *Intra*.
(P2) It is not permitted for any individual to defect in *Inter*.
(P3) There is no relevant difference between *Intra* and *Inter*.

These cannot all be true. *Which should go?*

Let me put my cards on the table. I will offer a defense of (P2). Basically, I will say that in *Inter*, it is unfair if you gas-guzzle (even before it is illegal). That is, it is unfair in the sense that *you* want people not to drive such cars, but still make an unjustified exception of yourself.

This has a surprising upshot. Possibly, (P3) should go and there must be some asymmetry between the two cases. But then what is the relevant difference? On the face of it, these cases look *very* similar. Indeed, as we will soon see, *Intra* and *Inter* share a similar structure.

Alternatively, (P1) should go and it is also problematic to relax on some particular day. Compare the following complaint: 'Why should I write today rather than my foregoing timeslices?' Perhaps we do not care as much about, let us call it, intrapersonal unfairness as we do about interpersonal unfairness. *Should we care more?*

I will proceed as follows. In §II, I describe the structure that *Intra* and *Inter* share. In §III, I summarize Chrisoula Andreou's important work on these dilemmas. In §IV, I show how Andreou's account supports (P1). In §V, I address the issue of unfairness in support of (P2). But if (P2) is true,

[1] For 'gas-guzzling' read: 'living a CO_2-intensive lifestyle'. This classic case is from Sinnott-Armstrong (2005), who instead used it to argue that it *is* permitted to defect in such cases, that is, as long as legislation is not in force—see §IV below.

and (P1) is supported by Andreou's account, then (P3) has to go. In §VI, I end the paper with a tentative suggestion on how (P3) may be denied.[2]

II. SYMMETRY

Where intrapersonal dilemmas involve an individual agent concerned with a certain long-term goal, interpersonal dilemmas are about multiple agents concerned with some collective goal. Examples of long-term goals are to not ruin your figure (prevent a bad outcome) or to write a book (realize a good outcome). Collective goals can be to not destroy the planet together (bad outcome) or to make it a better place (good outcome).

These cases share the same structure. Namely: they are characterized by a so-called 'preference loop' (Andreou 2006, 2007). Let us consider *Intra* first, which involves the following loop:[3]

> relax 0 days < relax 1 day < relax 2 days < relax 3 days < … < relax 50 days < relax 0 days

This loop is generated based on two preferences. First: I prefer to finish my book this year. Hence: I would rather write on too many days than relax on too many days. Second: I prefer to relax quite a bit. More specifically, I always prefer to relax for an extra day. This second preference may seem questionable. On day 50, for example, do I really prefer to relax for an extra day? It is right that I would prefer not to relax at all than to relax on too many days (as per preference 1). Even so, on day 50 (or day 360 for that matter), when I know that I am no longer able to meet my deadline, I still prefer to relax for an extra day (as I cannot meet it anyway).

A few important assumptions are in place. The effect of relaxing an extra day on the success of the project is taken to be negligible or trivial. You can always do that day's work later. Additionally, it is not plausible to think that there is a sharp threshold at which I will no longer be able to write the book if I cross it. If there *were* a sharp threshold n (for example 40 relaxation days), then the optimal strategy would be to relax on n–1 days (39). This would break the loop (I prefer to relax for 39 rather than 40 days). Instead, there is a vague threshold—between, say, 35 and 45

[2] This paper is restricted to cases like *Intra* and *Inter* where the dilemma is or at least can be solved, and I will bracket cases where no solution is expected (though see the appendix).

[3] These preferences are stable (do not change over time) and intransitive (A < B; B < C; C < A).

days—where it is unclear or indeterminate if I will be able to complete the book (and consequently there is no optimal strategy).[4]

Inter can be characterized by a similar loop:

> 0 persons gas-guzzle < 1 person gas-guzzles < 2 persons gas-guzzle < 3 persons gas-guzzle < ... < 50k persons gas-guzzle < 0 persons gas-guzzle

As before, this loop is generated based on two preferences. First: I prefer that the planet not be destroyed. Hence: I would rather have it that no one gas-guzzles at all (including myself) than that people gas-guzzle too much. Second: regardless of how many other people are gas-guzzling, I always prefer to gas-guzzle myself, that is to travel by car than to invest money and time in more sustainable ways of transportation (and so always prefer that one additional person gas-guzzles).[5] (If you do not like gas-guzzling, take some other behavior with a substantial carbon footprint: flying, eating meat, buying stuff, heating your home, etc. Alternatively, think of 'bigger' corporate-level decisions, such as to open further airports or build further oil drilling platforms, or to finance these activities—which could still be trivial as long as enough other parties make different choices.)

Again, the background assumption is that the effect of one extra ride with a gas-guzzler (or one extra seat in an airplane, etc.) on the state of the planet is negligible or trivial. A single ride with a gas-guzzler—or even an agent who gas-guzzles her whole life, or even one full gas-guzzling country—will not destroy the planet as long as enough others do not do that. Additionally, there is no sharp threshold at which the planet will be destroyed if we collectively cross it. As before, there is a vague threshold where it is unclear or indeterminate what will happen. In all these respects interpersonal dilemmas are similar to intrapersonal ones.

III. ANDREOU'S VIEW

The crux of Andreou's (2014) solution to these dilemmas lies in redescription. If certain conditions are met, individual acts can be redescribed as

[4] These assumptions seem plausible to me, and I will simply take them on board. See also Nefsky (2017, 2019).

[5] We can also ask what the group prefers, as opposed to an individual in it. Rather than saying that we always prefer an extra person to gas-guzzle, it may be more appropriate to say that we are collectively indifferent about how many people gas-guzzle, that is as long as the threshold will not be crossed. In §V, I discuss another interpretation of INTER based on fairness.

'doings in progress' that frustrate a certain goal of yours.[6] And if they can be redescribed in this way, they can be condemned as wrongful[7]—since you should not frustrate your own goals (an 'instrumental failure', if you like).

Consider the act of eating a brownie. Sometimes, Andreou says, this act can be redescribed as 'ruining one's figure', and this doing in progress frustrates one's long-term goal, namely, to keep one's figure. So, these could be two distinct descriptions of what I am doing:

(1) I am eating a brownie.
(2) I am ruining my figure.

When, exactly, can we redescribe (1) as (2)? We can do this when certain dispositions are in place, such that in due course I will have done (2). Dispositions "can involve fine-tuned habits of thought, appraisal, and action" (2014, 219) In this case, habits that ruin my figure are, for example, "to accept treats [I am] offered with the thought 'well, this one little treat will not make or break my chance of staying fashionably slim,' or without any thought at all and only the reaction 'yum'." (Andreou 2022, 37)

Three qualifications. First, (1) can be redescribed as (2) even if I do not intend to ruin my figure. After all, I do not have that intention, and yet I am doing it (given my dispositions). Second, (1) can be redescribed as (2) even if (2) has not been completed—my figure has not yet been ruined—when I am doing (1). I will be completing (2) only much later, after that given brownie. Third, (1) can be redescribed as (2) even if this doing in progress will, in fact, never be completed and it is interrupted at some point (unexpectedly). I might still be ruining my figure (before time t) even if I never get the chance to complete the ruination if, for example, brownies always happen to be sold out (after t).

When can we not redescribe (1) as (2)? (2) does not apply when relevant dispositions are not in place, or when I might have the relevant dispositions, but also a solution to halt the given doing in progress. If I am eating a brownie, but know it will be my last one, then I need not be ruining my figure. (*Intra* and *Inter* also describe examples of such solutions, and we will turn to them in §IV.)

[6] Or that achieve rather than frustrate them, that is, in the case of 'good' doings in progress (such as writing a book). Here, I will focus on 'bad' doings in progress.
[7] In light of the paradox, I will talk about 'wrongness' and 'impermissibility' rather than 'irrationality' throughout.

Just as individual acts can be redescribed as doings in progress that frustrate long-term goals of yours (such as to keep your figure), Andreou proposes, they can frustrate collective goals you have (such as protecting the planet). Consider:

(1) I am driving a gas-guzzler.
(2) We are destroying the planet.

(2) is an appropriate description of what I am doing, again, if certain dispositions are in place. This time, however, we should not simply consider the dispositions of a single agent, but of "several individuals combined" (2014, 219). Description (2) applies, then, when there is a group of people with certain dispositions such that in due course *we* will have done (2). If we all drive gas-guzzlers with the thought 'well, this one little ride will not destroy the planet', or even without any thought at all, then at some point we will have destroyed the planet.

As in the intrapersonal case, the idea is, (1) can be redescribed as (2) even if the interpersonal doing in progress (2) is not completed—the planet not destroyed—by my individual action alone.[8]

To summarize, sometimes individual acts can be redescribed as doings in progress that frustrate some goal of yours, long-term or collective. If that is so, what you are doing at some time "will not have only trivial effects" (2014, 218) and can be "seriously problematic relative to [your] concerns" (216). In this way, your action is wrongful, and you should not perform it. In steps (my reconstruction):

(1) I want to keep my figure.
(2) Hence, I should not ruin it. [1][9]
(3) My action of eating the brownie can be redescribed as 'I am ruining my figure'. [by Andreou's account]
(4) Hence, I should not eat the brownie. [2, 3][10]

[8] For complications regarding this step, see Supèr (2024). She argues that in interpersonal cases individuals cannot really be said to frustrate their concerns in the same way as they can frustrate their concerns in intrapersonal cases.
[9] Given a norm of coherence: if I desire state X, and me doing Y is necessary for X to occur, then I should do Y.
[10] Interestingly, the conclusion that you should not engage in self-torturing (in Quinn's 1990 thought experiment) can be reached in a similar way: when certain dispositions are in place, and a single choice of going one setting further (and accepting the money) can be redescribed as 'bringing yourself into a state of extreme pain'.

(1) I want the planet (and the life on it) to exist.[11]
(2) Hence, we should not destroy it. [1][12]
(3) My action of gas-guzzling can be redescribed as 'we are destroying the planet'. [by Andreou's account]
(4) Hence, I should not gas-guzzle. [2, 3]

There are some important differences with other outcome-oriented accounts (among others Nefsky 2017, Gunnemyr 2021). For Andreou, it is not about being a (tiny) cause of some bad outcome, nor about being a helpful cause of some better outcome. Firstly, it is about frustrating one's own desires (or 'concerns', as Andreou 2014 has it). Secondly, certain dispositions need to be in place, which produce an intrapersonal or interpersonal doing in progress (which in turn can frustrate one's desires).

INTERPERSONAL DOINGS IN PROGRESS

So far, so good. There is one complication in Andreou's account that I would like to address briefly. Namely: what is the range of actions that take part in a certain interpersonal doing in progress?

Imagine I accidentally drop a glob of glue on a stick that is lying on the ground. Others decide to make a kite with it. Can my action of dropping the glue be redescribed as 'we are making a kite'? Well, I am not acting with any shared intention to make a kite. But, on Andreou's account, such intentions are not required. We are not acting with any shared intention to destroy the planet and yet we are doing it (see also Kutz 2000, 186).

On Andreou's account, whether or not there exists a doing in progress depends on our dispositions. On the one hand, I am not part of the group that is making a kite, it seems, if we only consider *my* dispositions. I might be disposed to drop things accidentally, but this typically will not lead to the creation of interesting things. On the other hand, a kite is only to be expected given the intentions and dispositions of the others. They might even have the unusual plan to only use a stick on which some stranger accidentally dropped some glue.

One may think: who cares who is making the kite? Yet, we do care who is destroying the planet. If I am gas-guzzling, I am one of those who are destroying the planet (and acting wrongly). If I am only looking at an

[11] This is taken as a desire rather than an intention, as it is controversial if I can intend *that we protect the planet* (see Bratman 1999, ch. 8).
[12] Given: if I desire state X, and *us* doing Y is necessary for X to occur, then *we* should do Y.

advertisement for a gas-guzzler (say), I do not necessarily take part in the act of destruction (or act wrongly). This problem is not trivial. Consider Andreou here (in terms of Hardin's 1968 example):

> Suppose several herders change their ways and use the pasture at a fair and sustainable rate but most others do not, with the result that the pasture is still being destroyed. Are the herders that are using the pasture but showing restraint still part of the we that is destroying the commons? (2014, 223)

Andreou adds: "I see this as a gray area and I will leave the debate on the matter to those who have strong intuitions about the situation" (2014, 223). In response, I think the main problem is not so much that some have strong intuitions (that those who show restraint are not part of the group that is destroying the commons). Instead, the problem is that, if their action of 'showing restraint' can be redescribed as 'we are destroying the commons', then it follows (by reasoning similar to (1)-(4) above), that they are acting wrongly and should not show restraint or use the commons at all. That does not seem right.

Therefore, a further account to restrict the range of actions (that actually take part in some doing in progress, and so are problematic) would be helpful.[13] One possible account would be not to consider everyone's *actual* dispositions, but *hypothetical* ones. Would the (bad) outcome result if everyone (people or timeslices) acted on similar dispositions? If everyone were to think 'well, this single drive will not make a difference', we would indeed destroy the planet. But we would not if we all cared more and showed restraint.

Going hypothetical is, however, not something that Andreou would likely be attracted to. Even if I would destroy my figure if future timeslices acted on similar dispositions, it is not wrong for me to eat a brownie

[13] Andreou offers two brief replies. Her first point: even if those who show restraint do take part in the doing in progress, they might be excused. As an example: "even though someone is continuing the process of global warming by driving her child to the emergency room, she is not blameworthy because she is doing her part to reroute the harmful doing in progress by driving only when she really needs to." (2014, 223) In response to this, I agree that people may be excused (for acting wrongly) if they try their best 'to reroute the doing in progress with which they are involved'. Yet, I am not sure that we want to say that those who show restraint act wrongly in the first place. Andreou's second point: bystanders may be implicated in some doing in progress—by ignoring it and not trying to halt it—even when they do not take part in it. This, however, presupposes some separate account of when people take part.

now—she says, as we will soon see—as long as I have a solution to stop in time (as in *Intra*).

One more promising account would be to require that an action X only takes part in some doing in progress if the disposition on which one does X is 'relevantly involved' in the doing in progress (which is meant as a placeholder for some further story). My disposition of being careless, which causes me to drop the glue, is not relevantly involved in the creation of the kite. Similarly, then, dispositions on which people show restraint are also not relevantly involved in the destruction of the commons (see Björnsson 2021).

IV. PRECOMMITMENTS

Sometimes you have a solution to your intrapersonal dilemma. You have a solution because you installed a 'precommitment' to stop in time. You precommitted in the sense that you made it unattractive for your future timeslices to defect (see Andreou 2007).

Example: you want to quit smoking. In that case, you could make it physically harder to smoke (throw away your cigarettes, stop interacting with other smokers), or mentally harder (listen to podcasts on smoking and lung cancer), or impose financial penalties on yourself (as in *Intra*), or social costs (let your friends ignore you or make fun of you, as McClennen (1997, 234) imagines), or further penalties (for example deny yourself your weekly dinner out, as Andreou (2010, 206) does).

If defecting is sufficiently unattractive for your future timeslices, then the given doing in progress will be interrupted.[14] *Intra* is an example, and Andreou describes two more examples here:

> If we […] introduce, at some point along S's progression toward destroyed lungs, current or probable future adjustments to S's incentive and disposition structure so that there is now compelling reason to believe that, despite his taste for cigarettes, S will likely soon be quitting for good, then […] S's conduct as he opts to smoke the current cigarette may be well-advised. (2014, 215–216)

> There may be ways of doing this that are consistent with his continuing to eat the brownie. He may, for example, find a way to precommit to this brownie being his only brownie this year. But then he has

[14] Or continued, in the case of 'good' doings in progress. For example, you can book a writing retreat to make sure you will definitely focus and finish the manuscript.

> solved his problem and it is, other things equal, rational for him to eat this brownie, since, in this situation, his current doings really will have no more than a trivial impact in relation to his goal. (2022, 37)

A few things to note. First, precommitments are changes in your 'incentive structure' that *you* impose on yourself. Interventions by others (for example the government) might also work (that is, to halt the given doing in progress), but the idea here is that you do this in service of your own concerns.

In these cases, the precommitment is assumed to work. The agent has enough knowledge about herself and knows that her intervention will succeed in stopping the doing in progress. Of course, in real life, things could also be less certain. If you are not fully sure the precommitment will work, defecting may still be problematic.

Note that, in these passages, Andreou does not literally say that defecting (smoking one more cigarette, eating one more brownie) is permitted in these cases, but she says that it's 'well-advised' or 'rational'. Even so, I take the permission to be implied. Why would it be permitted to defect? This is permitted since doing so will not frustrate one's concerns. It is permitted to eat a brownie if you know that this will be the only brownie of the year and will not ruin your figure.

Are there analogous 'interpersonal' precommitments? Are there strategies we can impose on ourselves that make it unattractive for everyone to defect? As in the intrapersonal case, we could change the incentive structure and make it harder for ourselves to gas-guzzle. For example, we could lobby for carbon taxes that make gas-guzzling sufficiently unattractive, or even install complete bans (as in *Inter*).

REPLY TO SINNOTT-ARMSTRONG

Let us consider Sinnott-Armstrong's discussion of these strategies. As he argues, we should not pay high carbon taxes as long as they have not been introduced:

> We do not have any moral obligation to send a check to the government for the amount that we would have to pay if taxes were raised to the ideal level. (2005, 297)

With this, I agree. Collectively sending money will not necessarily protect the planet—unless there is some program in place that uses the money

for that purpose. If there is no such program yet, then it is pointless to send random checks (even collectively).

In contrast, collectively reducing one's carbon footprint *does* solve the problem, even without any government program in place (see Baatz and Voget-Kleschin 2019, 858; Davidson 2023, 688–689). Hence, why not stop gas-guzzling even before it is made illegal?

> If gas-guzzlers morally ought to be illegal, then maybe we morally ought to work to get them outlawed. But that still would not show that now, while they are legal, we have a moral obligation not to drive them just for fun on a sunny Sunday afternoon. (Sinnott-Armstrong 2005, 297)

So, Sinnott-Armstrong maintains, we may have a duty to lobby for stronger legislation, but we do not have any duty to stop gas-guzzling *before it is in force*. In *Inter*, then, it would be permitted to defect (before it is illegal).

I will readily grant that there are many cases where we should not adhere to a certain law before it is in force. To take a simple example: we should not drive on the left side of the road (if that will be a future law) when people currently drive on the right. In this case you do not want everyone to drive on the left side, or at least not before the law is in force.

In our case, in contrast, you do want no one to gas-guzzle. Specifically, you want no one to gas-guzzle *at all*, not just after it is enforced by law. After all, you know that if too many people do this, the planet will be destroyed. (Moreover, the problem can also be solved if everyone stops gas-guzzling—as well as further CO_2-intensive activities—even without any legislation.)

Of course, I am not sure if this is what Sinnott-Armstrong desires.[15] At any rate, this describes the agent I have in mind. This agent desires that the problem be solved, and hence that people do not gas-guzzle. If you do not care if the planet will be destroyed, then my story will not apply. Yet, it is important to see that the same goes for Andreou's approach. If you do not care whether the planet will be destroyed, then it will not matter if your action can be redescribed as a doing in progress that destroys it.

[15] At least, he wants the problem to be solved, and hence that governments "impose limits on emissions" and promote "alternatives to fossil fuels" (2003, 286).

Moreover, even if you do care, Andreou's account would render it permitted to defect in *Inter*. For, there is a solution to halt the doing in progress, and then the latter will not destroy the planet (and frustrate your concerns).

Yet, so long as you want people not to gas-guzzle, it still seems problematic to do this yourself. For, if you defect, you are *unfair*. You make an exception of yourself, not in that you allow yourself something illegal, but in that you allow yourself something you do not want others to do (regardless of whether it is illegal). If you agree, we need some further story.

You may think: collective efforts—international agreements, nationwide and corporate strategies—to halt climate change are far more *effective* than efforts by single individuals.[16] Again, I agree. Even so, this does not exclude individual actions still being wrongful. Furthermore, corporations and countries can act wrongly in much the same way—namely by being unfair. Specifically, as I will discuss next, it is unfair if you want *other* corporations and countries to cooperate to halt climate change, but make an exception of yourself for no good reason.

Note that I do not think that Andreou needs to deny that it is problematic to defect in *Inter*. Yet, her theory cannot account for this (in *Inter*, there is no instrumental failure, no case of frustrating one's own concerns) and so we need to look elsewhere.

V. UNFAIRNESS

Consider *Inter*, minus the precommitment solution. *Situation 1*: you know nothing about what others do. In this case, it is unfair to defect. You know it is not needed that everyone cooperates, but you also do not think people can make an exception of themselves for no good reason, and you do not have any such reason.

Situation 2: 100 people started to defect. Here, it is still unfair to defect. The explanation remains the same: you still want others not to defect.

Situation 3: you came together and, based on some democratic procedure, decided to change the incentive structure to not let it get out of hand. Namely: after 20k defections, defecting is punished harshly. How would this latter case be different? There is no reason why you (or anyone)

[16] Sinnott-Armstrong: "Global warming is such a large problem that it is not individuals who cause it or who need to fix it. Instead, governments need to fix it, and quickly" (2003, 203). See Johnson (2003) and Nihlén Fahlquist (2009).

may be among the 20k. The group does not say: it is fine to defect as long as the group of defectors does not get too big. The group simply says: we want to solve the problem, and in case some people want to be unfair, we have a solution. But that still implies that defecting is unfair.

The latter is how I intend to take *Inter*, and should be distinguished from another possible reading of the case. Namely, the group may also say: we *want* 20k people to defect. This group thinks it is great if some people gas-guzzle if that makes their lives better (and no other lives worse).[17] This group may give everyone a fair chance of buying a gas-guzzling car (for example on a first come, first served basis, or selection by lottery), and as soon as such a procedure is in place, it does not seem unfair for the lucky ones to defect. (The upcoming view will admit selected reasons to make an exception of oneself.)

Basically, there are two perspectives to consider: the group's perspective, and that of an individual in the group. The group may think: 'sure, some may gas-guzzle—as long as we have a solution to not let it get out of hand'. Individuals in that group, in contrast, might still consider it unfair if others defect.

These distinct perspectives correspond to two different moral concerns: one outcome-based, and one fairness-based. Andreou's account responds to the first concern. According to her proposal, to recall, it is wrong to defect when your individual action can be redescribed as a doing in progress that frustrates a long-term or collective concern of yours.

In the following, I will set forth a particular fairness account that responds to the second concern. Basically, this account says: it is wrong to defect when you want others to cooperate, but make an unjustified exception of yourself. Or, in terms of frustrating your concerns: you *would* frustrate them if others were to defect too. (This account is different from a more familiar fairness account, as I will soon explain.) Let us first see how this account would handle *Inter*:

[17] Andreou distinguishes between being partial to oneself and 'impartial benevolence' (2022, 8–9). If you are motivated by partiality, you care specifically about maximizing your own interests, and prefer to defect in interpersonal dilemmas for this reason. If you are impartially benevolent, you care about maximizing everyone's interests. As Andreou points out, interpersonal dilemmas arise even when everyone is impartially benevolent. In that case, you prefer that people defect so long as doing so promotes certain interests and does not undermine anyone else's interests. Note that in this case total well-being will be maximized only if people would not be jealous—which they may still be, even if they know the assignment was arbitrary.

(1) I want the planet not to be destroyed.
(2) Hence, I (should) want most people not to gas-guzzle. [1]
(3) Hence, gas-guzzling would amount to making an exception of myself. [2]
(4) This exception is unjustified.
(5) Hence, I should not gas-guzzle. [3, 4]

Note first that the argument has the very same starting point as (my reconstruction of) Andreou's view: the desire that the planet be protected. And, it has the exact same conclusion: that one should not gas-guzzle. Where they differ: in the middle. Basically, this argument has two parts: firstly, you want most people to act in some way (step to (2)), and secondly, you have no reason to make an exception of yourself (step to (5)).[18]

The step from (1) to (2) is fairly straightforward. If you are against some outcome X, you are against everything that is sufficient for X to occur. Given that widespread gas-guzzling destroys the planet, you do not want people to do this.

(Again: 'gas-guzzling' should be taken as a shorthand for living a CO_2-intensive lifestyle. You may still try to deny that this is sufficient for destroying the planet. Perhaps widespread gas-guzzling *plus* some technological solution that removes CO_2 from the atmosphere will not have that outcome. If that is a serious possibility—a big 'if'—then read (2) as a conditional claim: I want most people not to gas-guzzle on the condition that no solution has been implemented to extract enough CO_2 from the atmosphere.)

Note that the qualifier 'should' in (2) is not irrelevant. For, the desire in (1) also entails that one wants that most people do not fly, are vegan, etc. That is, one *should* want these things given what one wants in (1). Nevertheless, if you would ask people explicitly if they indeed want that, they may well deny that they have that desire—perhaps they do not know what protecting the planet requires or they fear that they, too, will have to stop flying or eating meat *on pain of unfairness*.

Importantly, it does not follow that you should want *everyone* to cooperate. That is not needed for the planet to be protected. It only follows

[18] Compare Mullins' two-step analysis, where you firstly believe that some situation is bad and that people should act in some way to fix it, and secondly that you should join them, based on an idea of impartiality—specifically, that you cannot respond with 'but my action makes no difference!'. See Benjamin Mullins, "The Inefficacy Argument and How to Respond to It."

that you should want *most* or *enough* people to cooperate, so that together we stay below the threshold of damage.

The step from (2) to (3) is only conceptual. If you want most others to do something, but not yourself, you make an exception of yourself (in this sense). The question is whether this is problematic. Hence (4): do you have any special justification? Is there any good reason why *others* should solve the problem *but not you*? Typically, in interpersonal dilemmas many reasons are not valid. That is, it is difficult to come up with reasons that do not also apply to numerous others. For example, you cannot say that you had a rough week and really need to relax in your gas-guzzler. Or that it is a great sunny day and that you should take advantage of it. Also, you cannot say, in *Inter*, that there is a solution to the problem, and hence that it does not matter if you defect. For, everyone else could appeal to that too.[19] If there is no valid reason, (5) follows: you should not make an unjustified exception of yourself, and so you should not gas-guzzle.

Of course, there are cases where (4) fails. Think of certain coordination cases. If you want to cross the street, you do not want others to pass at the same time. You want others to act differently, and make some sort of exception of yourself. Yet for good reason: you followed the traffic rules or saw it was safe. However, even in coordination cases (4) may hold. If you do not follow the traffic rules, or do not wait your turn, you might have no good reason why you should be exempted (but not others). In those cases, it could still be unfair to cross the street.[20]

Note that this argument makes no use of Andreou's redescription step (that is, of redescribing an individual act as a doing in progress). For, it is not about frustrating a concern of yours (which you can only do if your action can be redescribed in that way). Instead, it is about doing something—not gas-guzzling—that you also want others to do. Yet it is still *possible* to import Andreou's redescription step. Just like you do not want people to gas-guzzle, you do not want them to destroy the planet, and that is what they are doing (after redescription).

This point is also relevant in light of Nefsky's worry about fairness ideas in this context:

> Why does refraining from driving or flying count as *doing my part* in our preventing harmful climate change? […] Things will go the same

[19] One valid reason seems to be: others do not mind solving the problem for you (see Trifan 2020, 176–178). Yet, this does not seem to happen much in the real world.
[20] I analyze coordination cases in more detail in Wieland (forthcoming).

> with respect to the collective goal whether or not I do so. So, how can it count as doing a part of our achieving it? (2019, 4)

In response, the proposal here is not simply that we should all do our share. In that case, one might indeed ask: why refrain from flying (rather than, say, wear a protest T-shirt on the flight) if my conduct makes no difference? Instead, the idea is that we should not allow ourselves the sort of thing—flying in this case—that we do not want others to do.

TWO SORTS OF UNFAIRNESS

For sure, this is not the place to offer a full defense of this fairness argument. Instead, my aim here is to offer some support for (P2) of the paradox (see §I), that is, one explanation of why it is problematic to defect in *Inter*. Still, let me briefly say something about the account that underlies it, and distinguish this from another, more familiar proposal.[21]

A prominent fairness proposal is offered by Cullity (2000). According to Cullity, unfairness "involves arrogating to myself a privilege of not paying for benefits for which I rely on others' willingness to pay" (2000, 14). It is unfair, for example, to rely on others' willingness to protect the planet (and to benefit from that).[22] Cullity continues: "he relies on others to do what we ought collectively to be doing, without contributing himself […] he is leaving the work of meeting it to others" (15). Let us call this:

> Thick unfairness:
> There is a collective duty, and you let others do the work of meeting it.

As background, it is instructive to recall the controversy between Nozick (1974, 90–95) and Klosko (1987). According to Nozick, it cannot be unfair to let others do the work if you do not want them to do this work in the first place. In terms of one of his examples: "If each day a different person on your street sweeps the entire street, must you do so when your time comes? Even if you do not care that much about a clean street?" (1974, 94). The point: if you really do not want a clean street (or do not think the

[21] The upcoming distinction bears some similarities to the distinction between subjective and objective free riding (Bradley and Navin 2021). In the case of *subjective* free riding, you want some public good, and hence others to sustain it. In the case of *objective* free riding, these particular desires are not relevant: you simply benefit from some public good that you do not contribute to.

[22] I will abstract from this feature in this paper. I will take unfairness as letting others do the work rather than benefiting without returning the favor.

benefits are worth the costs of contributing to it), then it is not unfair if you decline to do your part.

To some extent, Klosko is sensitive to this worry: "there is a strong presumption that individuals should decide for themselves whether they are going to be required to make sacrifices" (1987, 247). Even so, according to him there are selected public goods—think of environmental protection and public health—to which everyone should contribute. These goods are, as he calls them, "presumptively beneficial", that is, they "can be presumed to be necessary for an acceptable life for all members of the community" (247). Or again: you should contribute to or pay for them *whether or not you also actually want them*.[23]

I am not going to take sides in this debate. For all I can see, in certain interpersonal dilemmas there could indeed be collective duties to solve them. However, if you are skeptical about these duties (such as for Nozick-style reasons), it is interesting to point out that my proposed account does not rely on them. The duty not to gas-guzzle did not rely on a collective duty to protect the planet, but rather on the agent's *own* desire that the planet not be destroyed. *You* want others to cooperate, but make an exception of yourself for no good reason. That is considered unfair, albeit in a different sense:[24]

> Thin unfairness:
> Regardless of whether there is any collective duty, you want others to do the work.

A main inspiration behind this account is Korsgaard's interpretation of Kant's formula of universal law (see Wieland 2024). This account specifically "appeals to [the] thwarting of the agent's *own* purpose" (Korsgaard 1985, 36).[25] To illustrate, consider Korsgaard's example of cheating on an entrance exam (42). In this case, I want the following things:

(1) I want to pass the entrance exam.
(2) Hence, I want many people not to cheat on it. [1]

[23] One may wonder: who would not want environmental protection or public health? Yet that is not the point. It is about when exactly people are being unfair.
[24] I call it 'thin' or 'minimal', as it does not assume any collective duty. In a different sense, you may think this type of unfairness is actually *worse*.
[25] That is: it evaluates whether the agent's 'purpose' or goal of her action is frustrated in a hypothetical world where others choose similar means, and not—à la Andreou—in the actual world after the redescription of one's action.

How does (2) follow from (1)? As Korsgaard explains, if many people were to cheat, then entrance exams would no longer be used as a criterion for selection. "Since a lot of incompetent people would get in, it would be found impracticable and some other method would be chosen" (42). Furthermore, if that is the case, I could not pass it in order to get in.[26] Here, too, it is not needed that everyone cooperates (refrains from cheating). Enough people should do it for such exams to remain in existence.[27] And yet, everyone should do it to avoid unfairness. Thus Korsgaard: "the test reveals unfairness, deception, and cheating" (36).

Note that this does not necessarily amount to a 'levelling down' type of fairness (Gunnemyr 2021, 36). The reasoning is not: 'I may not gas-guzzle, so no one may gas-guzzle' or 'They do not gas-guzzle, so I may not gas-guzzle', but rather: 'I want people not to gas-guzzle, and have no reason why I would be special in this regard'.

INTRAPERSONAL UNFAIRNESS

Interestingly, there is no initial reason to assume that this fairness account might not also apply to intrapersonal dilemmas. Just like it could be unfair to let other *people* do the work, it could be unfair to let other *timeslices* do the work. You want them to write, yet not yourself (current timeslice). If they were all to defect like you (relax and not write), then you would frustrate your long-term goal (finish your book). In steps:

(1) I want to finish my book.
(2) Hence, I want most timeslices not to relax. [1]
(3) Hence, relaxing now would amount to making an exception of my current timeslice. [2]
(4) This exception is unjustified.
(5) Hence, I should not relax now. [3, 4]

Is this similarly plausible?

VI. ASYMMETRY?

We previously described two perspectives: the group's perspective and the perspective of any individual in the group. On the group level, the primary concern is the outcome: it is fine if some people defect—as long

[26] See Roemer's test question: "What is the strategy I would like all of us to play?" (2019, viii). Also see Braham and van Hees (2020).
[27] The same applies to Kant's central case: enough (not necessarily all) people should refrain from making false promises in order for the practice to remain in existence.

as there is a solution to not let it get out of hand. The individuals may have additional fairness-based concerns. In the intrapersonal case, there are two similar perspectives: that of the diachronic agent (who exists through time) and the perspective of any timeslice of this individual. But, are the dynamics also fully analogous? In this final section, I will tentatively suggest a negative answer.

	Intrapersonal	*Interpersonal*
Perspective 1	Diachronic agent	Group
Perspective 2	Timeslice	Individual

A first question we may ask is: who is in charge? In intrapersonal cases, the diachronic agent can choose which timeslices can relax, and which not. One might think that this agent may even make such choices arbitrarily. In interpersonal cases, there is often no such authority. In *Inter*, there is for example no government that decides who may gas-guzzle and who not. Moreover, even if there were such an authority, we would not accept them arbitrarily favoring certain people over others.

However, this does not really seem to explain any asymmetry. Even if we, in fact, let the diachronic agent arbitrarily favor certain timeslices over others, it does not follow that this is how it should be. *Why* may this agent do this?

I think we should ask: *who exactly cares about what?* In intrapersonal cases, a certain timeslice, one might think, need not care much about whether or not some long-term goal will be met. After all, the achievement of the goal will only affect *future* timeslices. What does my current timeslice care—following this thought—if I finish my book when it will no longer exist at that point?

In interpersonal cases, in contrast, a certain individual in the group does care about the outcome (for example whether or not the planet will be destroyed). This outcome will affect her too, not just the group. (Interpersonal dilemmas could also have a diachronic dimension, where the outcome will only affect future members of the group (Gardiner 2002). We will bracket these mixed cases here, and only compare 'pure' intrapersonal and interpersonal dilemmas that involve long-term and collective concerns respectively, not both.)

What follows? If timeslices have no outcome-based concerns, they seem to have all the more reason to file fairness complaints. For, if the outcome does not matter to them (insofar as it will not affect them), they

do not necessarily want other timeslices to cooperate, and they are not being unfair (at least not in the thin sense discussed) if they defect. For that reason, they may complain if they are not allowed to defect (relax).

However, there is some reason to believe that their perspective and concerns do not matter too much. It is the following: in intrapersonal dilemmas, the diachronic agent is frustrating a concern of hers (not necessarily of her timeslices). The agent that exists through time relaxes too often and fails to complete her book. In interpersonal dilemmas, in contrast, the group is not only frustrating her own concerns (if she has them at all), *but especially the concerns of the individuals in it*.

In other words, in intrapersonal dilemmas perspective 1 is leading (the diachronic agent), while in interpersonal dilemmas perspective 2 is leading (the individuals in the group). That is quite a difference.

	Intrapersonal	*Interpersonal*
Perspective 1	**Diachronic agent**	Group
Perspective 2	Timeslice	**Individual**

Now if interpersonal dilemmas are mostly about the concerns of the individuals (and intrapersonal dilemmas are *not* about the concerns of the timeslices), then *that* might explain—perhaps—why fairness issues are more pressing in the former. This asymmetry merits further attention.[28]

APPENDIX

This paper addressed cases like *Intra* and *Inter* where the dilemma is or at least can be solved. Sometimes, though, there is no solution in sight. Sometimes I rationally expect most people to defect, because I believe that *they* believe that defecting will benefit them individually, and make little difference to the planet's destruction. Suppose I decide to defect because I expect others to defect (for the reason just given). In that case, we might think: I do not make an exception of myself. After all, I am simply doing what I expect others will be doing. Yet this presents a challenge to the proposed account. I said: it is unfair if I want others to cooperate but make an exception of myself. That still holds in this case: I want others to cooperate—it is just that I believe they will not.[29]

[28] That is, by fairness theorists specifically, and everyone interested in a *uniform* account of intra- and interpersonal dilemmas more generally.
[29] I discuss this issue—raised by the reviewer—in terms of interpersonal dilemmas, but, again, a similar analysis may apply to intrapersonal dilemmas.

Two points in response. First, I should not rationally expect others to defect in the first place. Even if others believe that defecting will benefit them individually, and make little difference to the planet's destruction—all true—they may still cooperate. For, even when they believe defecting will benefit them, they may still *also* believe that doing so would be unfair. In that case, I should not expect them to defect, and I should not defect for that reason. Of course, I cannot be *assured* that they will cooperate, but this does not mean that I should expect them to defect (unless I receive further information).

Sometimes, though, we have more information. Perhaps I know that the others do not care about unfairness, and so are likely to defect. Or perhaps they do care about it to some extent, but they cannot be assured that enough people will cooperate, and have a tendency to defect for that reason. If this is what I believe, it can be rational for me to expect that they defect. If I then decide to defect, I do not seem to make an exception of myself. I am not only doing what I expect others will be doing, but I am even acting for similar reasons. I defect because I cannot be assured that enough people will cooperate.

It is right that, in such a case, I do not make an exception of myself in the sense that others (are expected to) act differently (or that I act for reasons that others do not or cannot act on). Even so, I *do* make an exception of myself in the 'thick' and 'thin' senses discussed in this paper: I do not do my share in discharging a collective duty to solve the problem, and I still want others to solve it and lack a good reason to single myself out. In these ways, I can still be considered unfair. If we all decide to defect, *we* are all unfair—not relative to what others are doing (who also defect), but relative to what we all should be doing or to what I want us to be doing. That's the proposal.[30]

REFERENCES

Andreou, Chrisoula. 2006. "Environmental Damage and the Puzzle of the Self-Torturer." *Philosophy and Public Affairs* 34: 95–108.

Andreou, Chrisoula. 2007. "Environmental Preservation and Second-Order Procrastination." *Philosophy and Public Affairs* 35: 233–348.

[30] In Wieland (2024, §6), I offer an alternative analysis: if too many people are defecting, I could suppress my desire that they cooperate. After all, it is an idle hope. I then defect only because it is no longer possible to solve the problem. Moreover, if I drop my desire—premise (1) in the arguments discussed in this paper—it no longer follows that I should cooperate.

Andreou, Chrisoula. 2010. "Coping with Procrastination." In *The Thief of Time*, edited by Chrisoula Andreou and Mark D. White, 106–115. Oxford: Oxford University Press.
Andreou, Chrisoula. 2014. "The Good, the Bad, and the Trivial." *Philosophical Studies* 169: 209–225.
Andreou, Chrisoula. 2022. *Commitment and Resoluteness in Rational Choice*. Cambridge: Cambridge University Press.
Baatz, Christian, and Lieske Voget-Kleschin. 2019. "Individuals' Contributions to Harmful Climate Change: The Fair Share Argument Restated." *Journal of Agricultural and Environmental Ethics* 32: 569–590.
Björnsson, Gunnar. 2021. "Being Implicated: On the Fittingness of Guilt and Indignation over Outcomes." *Philosophical Studies* 178: 1–18.
Bradley, Ethan, and Mark Navin. 2021. "Vaccine Refusal is not Free Riding." *Erasmus Journal for Philosophy and Economics* 14: 167–181.
Braham, Matthew, and Martin van Hees. 2020. "*Kantian* Kantian Optimization." *Erasmus Journal for Philosophy and Economics* 13: 30–42.
Bratman, Michael. 1999. *Faces of Intention*. Cambridge: Cambridge University Press.
Cullity, Garrett. 2000. "Pooled Beneficence." In *Imperceptible Harms and Benefits*, edited by Michael Almeida, 1–23. Dordrecht: Kluwer.
Davidson, Marc. 2023. "Individual Responsibility to Reduce Greenhouse Gas Emissions from a Kantian Deontological Perspective." *Environmental Values* 32: 683–699.
Gardiner, Stephen. 2002. "The Real Tragedy of the Commons." *Philosophy and Public Affairs* 30: 388–416.
Gunnemyr, Mattias. 2021. "Reasons, Blame, and Collective Harms." PhD dissertation, Lund University.
Hardin, Garrett. 1968. "The Tragedy of the Commons." *Science* 162: 1243–1248.
Johnson, Baylor. 2003. "Ethical Obligations in a Tragedy of the Commons." *Environmental Values* 12: 271–287.
Klosko, George. 1987. "Presumptive Benefit, Fairness, and Political Obligation." *Philosophy and Public Affairs* 16: 241–259.
Korsgaard, Christine. 1985. "Kant's Formula of Universal Law." *Pacific Philosophical Quarterly* 66: 24–47.
Kutz, Christopher. 2000. *Complicity. Ethics and Law for a Collective Age*. Cambridge: Cambridge University Press.
McClennen, Edward. 1997. "Pragmatic Rationality and Rules." *Philosophy and Public Affairs* 26: 210–258.
Nefsky, Julia. 2017. "How You Can Help, Without Making a Difference." *Philosophical Studies* 174: 2743–2767.
Nefsky, Julia. 2019. "Collective Harm and the Inefficacy Problem." *Philosophy Compass* 14: 1–17.
Nihlén Fahlquist, Jessica. 2009. "Moral Responsibility for Environmental Problems: Individual or Institutional?" *Journal of Agricultural and Environmental Ethics* 22: 109–124.
Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.
Quinn, Warren. 1990. "The Puzzle of the Self-Torturer." *Philosophical Studies* 59: 79–90.
Roemer, John. 2019. *How We Cooperate. A Theory of Kantian Optimization*. New Haven, CT: Yale University Press.

Sinnott-Armstrong, Walter. 2005. "It's Not My Fault: Global Warming and Individual Moral Obligations." *Perspectives on Climate Change* 5: 221–253.

Supèr, Tessa. 2024. "Collective Doings in Progress and the Attribution Problem." *Erasmus Journal for Philosophy and Economics*, 17 (1): 149–168.

Trifan, Isabella. 2020. "What Makes Free Riding Wrongful? The Shared Preference View of Fair Play." *Journal of Political Philosophy* 28: 158–180.

Wieland, Jan Willem. 2024. "Cooperation – Kantian-Style." *Inquiry*, 1–26.

Wieland, Jan Willem. Forthcoming. "Kantian Free Riding." *Journal of Ethics and Social Philosophy.*

Jan Willem Wieland is an Associate Professor of Ethics at Vrije Universiteit Amsterdam, specializing in collective action and the ethics of inefficacy. Contact e-mail: <j.j.w.wieland@vu.nl>