

# Collective Doings in Progress and the Attribution Problem

TESSA SUPÈR

*Vrije Universiteit Amsterdam*

**Abstract:** We often encounter situations in which an undesirable outcome is brought about through a series or collection of seemingly inconsequential actions. This phenomenon, referred to as the inefficacy paradox, occurs both intrapersonally and collectively. Paradoxically, while we have good reason to avoid such patterns of action, there appears to be no compelling reason to abstain from any of the individual actions constituting such a pattern given its trivial impact. This paper scrutinizes Chrisoula Andreou's prominent endeavor to resolve the inefficacy paradox in both the intrapersonal and the collective context by utilizing their structural similarity. While her approach may prove successful in intrapersonal cases, the applicability of the proposed solution to collective cases is, I argue, ultimately limited. This is due to a fundamental dissimilarity between these two kinds of cases, which I lay out in the paper. This insight may also shed light on the transferability of other proposed solutions to the inefficacy paradox from the intrapersonal to the collective context.

**Keywords:** trivial differences, collective doings, temporally extended doings, inefficacy paradox, action attribution

**JEL Classification:** Q54, Q56, D81

## I. INTRODUCTION

In everyday life, we are often confronted with situations in which a series or collection of actions jointly leads to an undesirable or even destructive outcome while each action individually makes no more than a trivial difference to that outcome. This pattern, which can be referred to as the *inefficacy paradox*, occurs both intrapersonally, for example, when I fail to save for retirement or procrastinate studying for an exam, as well as collectively, for instance, when we together damage the environment due to high-emission lifestyles or cause traffic jams by failing to carpool. The predicaments that the inefficacy paradox leads to in such *trivial*

*difference cases* can be seen as the main motivation for seeking to resolve the paradox. Given the structural parallel between intrapersonal and collective cases, some authors have attempted to extend potential ways to resolve the inefficacy paradox from one domain to the other. In this paper, I will focus on one such prominent attempt by Chrisoula Andreou, who defends the position that what gives rise to the inefficacy paradox relies on a misguided idea of what is going on in trivial difference cases, and that an alternative characterization of these cases resolves the paradox in both intrapersonal and collective cases. I argue that while Andreou's approach is promising in the intrapersonal context, it ultimately does not transfer to the collective context due to a fundamental difference between intrapersonal and collective trivial difference cases that has thus far been left unaddressed.

## II. TRIVIAL DIFFERENCE CASES AND THE INEFFICACY PARADOX

Let's start by considering an example of a trivial difference case, adapted from Andreou (2006; 2014), which illustrates how the inefficacy paradox can manifest in the intrapersonal context:

### *Smoker*

Consider an agent, S, who enjoys smoking several cigarettes each day. Being perfectly aware of the health risks associated with smoking, S knows that if she were to continue doing so for a long enough time, she would eventually end up with irreversibly damaged lungs.<sup>1</sup> She cares greatly about her pulmonary health and is concerned with maintaining her lungs in good condition. At the same time, S rightly believes that smoking one cigarette has such a minimal effect on her lungs, that no matter in what stage of lung damage she currently finds herself, it would not make a difference if she were to smoke one more cigarette.

Let's next consider a different case that Andreou (2014) presents (adapted here) in order to show how the inefficacy paradox can also arise in a collective context:<sup>2</sup>

---

<sup>1</sup> For the sake of illustration, I assume here that continued smoking inevitably leads to irreversible lung damage. There may, of course, be interpersonal differences in the effect that smoking cigarettes actually has on one's pulmonary health.

<sup>2</sup> The case is an adaptation of the classic *tragedy of the commons* case, most famously discussed by Hardin (1968).

*Herder*

Consider an agent, H, who makes use of a common pasture to graze his herd. The herd that he keeps is very large, and H is aware that if too many herders use the common pasture to graze a herd of the same size, the pasture will be damaged irreversibly. Since his livelihood depends on access to the pasture, H is concerned with maintaining it in good condition. At the same time, H rightly believes that grazing just his own herd (albeit oversized) has only minimal effect on the state of the pasture, so that, no matter what its current condition is, it would not make a difference if he were to graze his oversized herd in the common pasture.

The structure of these cases presenting the inefficacy paradox is analogous: in both the intrapersonal and the collective case, there is an outcome that unfolds through the effects of a large set of actions, but given the trivial effect that *each* action brings about, none of these actions individually makes a difference to the occurrence of the outcome.

It is important to distinguish between what I consider to be two different (though related) problems that the inefficacy paradox gives rise to in such cases. On the one hand, there is a *theoretical* problem: as Andreou (2014) notes, the structure of cases like *Smoker* and *Herder* raises “complications concerning the assessment of conduct” (211). We tend to assess bits of conduct (or, doings) in relation to (a set of) relevant concerns and the extent to which these concerns are affected by what is being done. For example, ending up with irreversibly damaged lungs would be an unacceptable outcome relative to S’s concern with maintaining healthy lungs.<sup>3</sup> Assuming that continuously smoking cigarettes throughout her life would significantly affect her lungs, this would be considered unacceptable. Yet, given the trivial impact that each cigarette has on her lung health, it seems that each time S smokes a cigarette, nothing unacceptable is happening relative to the concern at hand. Similarly, given the (shared) concern that each herder has with maintaining the viability of the common

---

<sup>3</sup> Following Andreou (2014), I assume throughout this paper that the acceptability of an outcome is solely determined by the extent to which that outcome would frustrate a relevant concern. The reason *why* frustrating a particular concern is unacceptable (and thus, why a certain outcome would be considered unacceptable) can differ, depending on the nature of the concern. For example, if it is a morally relevant concern, frustrating this concern would be morally unacceptable, whereas if it is a prudentially relevant concern, the outcome would be prudentially unacceptable. I will not be concerned with this issue here. For the present purposes, what matters is that *if* an agent has a concern, any outcome that conflicts with it would be unacceptable.

pasture, ending up with an irreversibly damaged pasture would be an unacceptable outcome. Assuming that if all herders grazed their oversized herds on the common pasture, it would not be sustained, this would be considered unacceptable. Yet, it is also assumed that the impact of each individual herder's (over)use of the pasture is trivial, and so it seems that what is going on when each herder grazes his herd is acceptable. In both types of cases, we seem forced to draw the paradoxical conclusion that something unacceptable is unfolding and yet there seems to be no point along the way at which anything unacceptable is going on. To the extent that this puzzling assessment of conduct is considered undesirable, the theoretical problem warrants a solution that would allow for a comprehensible assessment of what is going on in trivial difference cases.

However, insofar as these assessments are conceived of as merely descriptive, it could be argued that the theoretical problem is not particularly problematic *per se*. Indeed, it is mainly when we recognize the normative force the paradoxical conclusion has in an agent's practical decision-making, that the real challenge presents itself. The (un)acceptability of what is going on when an agent engages in a particular behavior can determine whether it would be acceptable for her to do so (though, as we shall see, not in every case). And to the extent that we require agents to base their practical deliberations about what to do based on what would be acceptable, the inefficacy paradox gives rise to a practical conflict: each time S contemplates smoking a cigarette, or any of the herders considers grazing their oversized herd on the common pasture, they each realize that it would be acceptable to do so, and thus find no reason not to.<sup>4</sup> As such, they are forced to go down a self-destructive pattern of conduct leading to an outcome that, at the same time, they would be forced to avoid given their concerns. Hence, we may also formulate a *practical* problem: while there is reason to avoid the set of acts that leads to the unacceptable outcome, there seems to be no reason to refrain from any of the individuals acts that, together, cause that outcome. So, while the

---

<sup>4</sup> Given that we assumed that the acceptability of an action depends on whether that action contributes in a significant way to the outcome in question, the reasons I am concerned with here are of instrumental nature. That is not to say that there cannot be other reasons— independent of the outcome —to refrain from an action, such as (in the context of *Smoker*) the steep price of cigarettes or a promise made to one's parents, or (in the context of *Herder*) the proximity of another, equally good pasture or a law that prohibits overgrazing. Such reasons, however, are inherently contingent on case-specific details and do not help us in solving the more fundamental problem that is at stake in trivial difference cases, and will therefore be disregarded here (this approach is similar to Nefksy (2019)). Any mention of *reasons* throughout this paper should therefore be read as reasons that directly and instrumentally derive from the outcome in question.

theoretical problem concerns the evaluation of conduct, the practical problem concerns reasons for action. The practical problem is arguably the most important rationale underlying attempts to resolve the inefficacy paradox (for example see Sinnott-Armstrong 2010; Tenenbaum and Raffman 2012; Nefsky 2019; Soon 2021). A satisfying solution to the inefficacy paradox should therefore settle not only the theoretical problem that it gives rise to, but also—and especially—the practical problem: it should provide a reason for the agent not to engage in the (individually trivial) actions that ultimately lead to the unacceptable outcome. Importantly, however, as will become apparent, a solution to the theoretical problem does not automatically function as a solution for the practical part. In what follows, I will discuss Andreou’s approach to resolving the inefficacy paradox, and the extent to which her approach constitutes a solution to the problems presented above.

### III. DOINGS IN PROGRESS

In describing what is happening in the trivial difference cases at hand, we have so far been saying that each time S smokes a cigarette, or H grazes his herd, what is going on does not, in any perceptible way, make a difference to the outcome that would frustrate their relevant concerns. Call this the *standard view*. Andreou (2014) argues that the standard view relies on a misguided conception of what is going on in trivial difference cases. An accurate way of characterizing what is going on in such cases will, according to Andreou, lead to a rejection of the idea that, each time one of the above-mentioned doings is performed, *what is going on* has only a trivial effect on the outcome in question. Call this the *alternative view*.

To understand this better, it is important to highlight some background ideas. First, doings are typically temporally extended, in the sense that they “unfold over time” (Andreou 2022, 34). That is, even the smallest doings—for instance, in the context of S’s smoking, inhaling a bit of smoke from a cigarette—are constituted by even smaller doings (raising the cigarette up to one’s lips, sucking on it, and so on) which require at least some time to be performed. Andreou (2014, 2022) refers to such temporally extended doings as *doings in progress*.

This leads to the second point, namely that for any point in time, there are multiple ways to characterize what is currently going on. For example, take a certain moment  $t^*$ , during which S is drawing on her cigarette. While ‘inhaling a bit of smoke’ would be an accurate answer to the

question what is going on at  $t^*$ , descriptions such as ‘smoking a cigarette’ or ‘taking a break from work’ would qualify as equally accurate, even though the latter two doings *extend beyond*  $t^*$ . This leads to the observation (related to the work of, for example, Anscombe 1963 and Thompson 2008) that what is going on *at* a particular moment in time “need not be the sort of thing that can fit *in* this moment” (Andreou 2014, 215).

Importantly, Andreou (2014, 2022) adds a third idea, namely that the set of doings that are in progress at a certain time is not limited to *intentional* doings. For instance, say that at  $t^*$ , the smoke from S’s cigarette is blowing in the face of a passerby, who happens to detest the smell of cigarette smoke. Although S merely intended to enjoy her cigarette, what is going on at  $t^*$  could now also accurately be characterized as ‘bothering a stranger.’

The crucial point, then, which Andreou (2014) makes based on these three ideas is that “given the right conditions, where this includes facts about S’s dispositions” (215), we can add another, equally accurate description of what is going on at  $t^*$ , namely that at  $t^*$ , S is ‘irreversibly damaging her lungs.’ Obviously, irreversibly damaging her lungs has, contrary to smoking one cigarette, much more than a trivial impact on S’s pulmonary health, and thus frustrates her concern.

Before we can appreciate the implications of Andreou’s alternative view in cases like *Herder*, one additional background idea needs to be laid out which is relevant in the context of collectively caused outcomes. Andreou (2014) writes that

we talk not only about individuals doing things, but also about groups of individuals doing things, where these things might be such that they could not or just would not be accomplished by any individual in the group working solo. (217)

The point made here is that while doings in progress involve multiple actions performed over time, these actions need not always be performed by a single agent; there are also many cases in which there is a *collective* doing in progress, involving a series or collection of actions performed over time by *multiple* agents. However, as Andreou notes, it still holds true that such a (collective) doing can be *in progress at* a moment in time without it being *completed in* that moment. Furthermore, to say that this collective doing is in progress, it need not even be true that all agents involved in the doing are currently, in that moment, actively engaged in

the doing. For example, when two people are organizing a dinner party, and one is inside setting the table for the guests, while the other is standing outside taking a quick phone call from his mother, it would still be accurate to describe what is going on at that moment in terms of the collective doing ‘organizing a dinner party’ even though *in* that moment only one of the organizers is actively engaged in the doing.

Going back to *Herder*, then, and looking at a moment in time  $t^*$ , during which H is grazing his herd, Andreou argues that, assuming the relevant facts about the world and dispositions of the other herders are in place, describing what is going on at  $t^*$  either as ‘grazing a herd in the common pasture’ or as ‘irreversibly damaging the common pasture’ would be equally accurate. While the former does not (per assumption) in any significant way affect the relevant concern in question (that is to say maintaining a viable common pasture), the latter obviously does.

To summarize, Andreou’s approach to resolving the inefficacy paradox in cases like *Smoker* and *Herder*, is to show that the paradox arises only if we fail to acknowledge that what is going on *in* a moment in time is not *all* that is going on. In other words, descriptions like ‘smoking a cigarette’ and ‘grazing a herd in the common pasture’ are accurate, but not exhaustive when it comes to describing what is going on at  $t^*$ . Looking only at doings that are completed *in*  $t^*$  would simply give us an incomplete picture of what is going on.

Much more can be said about this way of depicting trivial difference cases. For instance, Andreou (2014) herself goes into discussing some potential objections to her approach, and Tenenbaum (2023) questions the extent to which *all* relevant doings can be modeled in the way presented here. I will not go into these worries here: my aim is not to challenge the (plausible) claim that a description of what is going on at any moment in time can include things that extend beyond that moment, nor that it can include the actions of multiple agents. Instead, my aim for the remainder of this paper will be to argue that there is only a limited extent to which Andreou’s alternative view of what is going on in trivial difference cases constitutes a solution to the problems that, under the standard view, arise in these cases.

Starting with the theoretical problem, it should be relatively straightforward to see how Andreou’s alternative view of trivial difference cases allows us to reject the paradoxical assessment that something unacceptable is going on over time, while within that time, there seems to be no point at which anything unacceptable was happening. To evaluate the

acceptability of what is going on at a certain point in time, we must consider *everything* that is going on. Importantly, as we have seen with the alternative view, this can include doings that extend beyond that moment and, in collective cases, beyond the agency of a single individual. In cases like *Smoker* and *Herder*, this led to the conclusion that at each moment when a cigarette is being smoked, or a herd is being grazed, what is going on includes—in addition to the trivial, momentary doings—doings that *do* frustrate the relevant concern in question. So, it can be concluded that what is going on *at* those moments is not acceptable, despite the trivial effects of what is going on *in* those moments. This way, the alternative view seems to give us a way out of the theoretical problem: on Andreou's view, what is going on at each moment along the way to an unacceptable outcome is *not* always acceptable, so that, in contrast to what was originally suggested under the standard view, there is no need to draw a paradoxical conclusion regarding the assessment of conduct in trivial difference cases.

However, recall the desideratum that a satisfying solution to the inefficacy paradox should also be able to settle the practical problem that the paradox gives rise to. To explore if Andreou's alternative view can indeed do so, the next question that should be addressed is: are we now in a position to conclude that, if S were to smoke a cigarette, or if H were to graze his herd in the common pasture, *they* would be doing anything that frustrates their respective concerns? If so, it could be concluded that if they were to perform these actions, they would be acting unacceptably, which would give them a reason to refrain from doing so. In order to answer this question, we need to understand when a description of *what is going on* at a particular time is relevant for assessing what *an agent is doing* at that time. Andreou herself does not explore this question in detail. As such, it is unclear whether her view can indeed be of use for solving the practical predicament that agents in trivial difference cases find themselves in. I will address this issue in what follows, which will eventually lead me to challenge the idea that the extended descriptions of what is going on are always relevant for assessing the acceptability of the relevant agent's conduct. This, in turn, will lead me to reject the idea that the alternative view can, in an important category of cases, solve the practical problem that the inefficacy paradox gives rise to.



#### IV. THE ATTRIBUTION PROBLEM

Let's start by distinguishing between what is *going on* and what is *being done* at a moment in time. Consider W, who is walking to work and is currently, at  $t^*$ , crossing the street to get to his office. When asked the question what is *going on* at  $t^*$ , we have seen this could accurately be answered with: 'W is crossing the street.' Likewise, when asked what is *being done* at  $t^*$ , the same answer seems appropriate. But say that, at  $t^*$ , it also happens to be raining. In that case, the former question could, alternatively, be answered with: 'it is raining,' while this would *not* be an appropriate answer to the latter question. There is an obvious difference between these two descriptions: while it is true that both 'W is crossing the street' and 'it is raining' are *in progress* at  $t^*$ , only the former qualifies as a *doing* that is in progress. So, it seems, there is a relevant difference between describing what is *going on* and what is *being done* at a moment in time: the latter is limited to doings, while the former may also include descriptions of other states of affairs. Note that descriptions about other states of affairs *can*, in many cases, certainly influence how we assess an agent's conduct. For example, if W were to be strongly concerned with not arriving at work with wet clothes, then W would be criticisable for walking to work without taking any measures that would prevent him from getting rained on. However, since 'it is raining' is not a description of anything that is being *done*, it follows that 'it is raining,' by itself, is not something that W could be criticized for. So, if a description of what is *going on* at  $t^*$  is to say anything (in the relevant sense) about the acceptability of what an agent is doing at  $t^*$ , it must minimally be the case that it is a description of something that is *being done*.

Note that I say minimally. Indeed, not just *any* description of what is being done at  $t^*$  can be used to evaluate an agent's behavior at  $t^*$ . Recall that, following Andreou, descriptions of what is being done at  $t^*$  can include the actions of multiple agents. Say that at  $t^*$ , while W is crossing the street, another agent D is driving her car on the road that W is crossing. Now, what is being done at  $t^*$  could be described as 'D is driving her car.' Even though this description, unlike 'it is raining,' does qualify as a *doing* that is in progress at  $t^*$ , it cannot, in any relevant way, factor into the evaluation of what W is doing at  $t^*$ . If we compare the description 'W is crossing the street' and 'D is driving her car,' the difference, again, is obvious: the former seems to be a description of something that is being done *by W*, while the latter does not. So, what follows is the conjecture that a description of *what is going on* at a particular time is relevant for

evaluating what *an agent is doing* at that time only if it is a description of something that is being done *by the agent* in question. In the examples discussed so far in this section, it may be difficult to see the relevance of spelling out this conjecture: it seems self-evident that descriptions like ‘it is raining’ and ‘D is driving her car’ do not influence (in the relevant sense) how we evaluate W’s behavior at  $t^*$ . It shall soon become apparent, however, that in the other cases of interest here, things are not as obvious.

In the previous section, we established that in both *Smoker* and *Herder*, there is something unacceptable *going on* (relative to the relevant concern of the agent in question). Currently, we are exploring if we can take this to mean that the agent in question is *acting* unacceptably. Given the conjecture sketched above, we must thus investigate if the descriptions of what is *going on* indeed refer to something unacceptable that is *being done by the relevant agent*. A first step in this inquiry would be to see if the descriptions refer to a doing that can be *attributed to* the relevant agent. For, as far as I can tell, it would seem uncontroversial to state that an agent can only be described as performing a certain doing if that doing can be attributed to her. Much has been said about the issue of attributing doings to agents (for example see Wolf 1990; Watson 1996; Korsgaard 2008; Velleman 2015), and I am by no means concerned here with settling any dispute on this topic, or providing an exhaustive set of criteria for attribution. Instead, what I am concerned with is how, according to Andreou (2022), the temporal extendedness of a doing affects its attributability:

Given that temporally extended actions require temporally extended agents, it is misguided to think about what was going on [at  $t^*$ ] as executed by a non-temporally extended agent [...] Being temporally extended, any doing going on at  $t^*$  must be attributable to an agent that endures for more than that single point in time. [...] Once it is recognized that a doing at  $t^*$  will normally reach beyond  $t^*$  and be attributable to an entity sufficiently temporally extended to carry out the doing, it becomes clear that, as with figuring out what doings are going on at  $t^*$ , figuring out to whom to attribute the doings at  $t^*$  is not a matter of looking for entities in  $t^*$ . (38)

What should be taken from this passage? According to Andreou, a temporally extended doing, which extends *beyond*  $t^*$ , can only be attributed to an (at least) equally temporally extended agent, and not to a ‘time-slice’

agent, who persist only *in*  $t^*$ . What does this imply for the possibility of attributing the doings to the relevant agents in *Smoker* and *Herder*?

Let's start with *Smoker*. We have seen that what is going on at  $t^*$ , when S smokes a cigarette, can be accurately described as 'irreversibly damaging her lungs' (which is unacceptable, in the sense that it frustrates a relevant concern of S). It seems clear that this is a description of a *doing*, but is it also true that this doing can be attributed to the relevant agent? Note first that in *Smoker*, we can distinguish between two "levels of agency" (Andreou 2014, 217): there is S, who is a temporally extended agent in the sense that she persists *beyond*  $t^*$ , and there is  $S^*$ , a time slice of S, who, unlike S, persists only for as long as  $t^*$ .<sup>5</sup> Next, it should be noted that 'irreversibly damaging her lungs' is a doing that extends temporally beyond  $t^*$ . According to Andreou's conceptions about the attribution of temporally extended doings then, this doing cannot be attributed to a time-slice agent but requires instead an agent who is also temporally extended. It would therefore have to be concluded that, while 'irreversibly damaging her lungs' *can* be attributed to the temporally extended agent S, it *cannot* be attributed to the time-slice agent  $S^*$ . In other words, in *Smoker*, we can accurately describe what is going on at  $t^*$  as 'S is irreversibly damaging her lungs,' but the constraint of attributing temporally extended doings only to temporally extended agents, limits us from saying that at  $t^*$ , ' $S^*$  is irreversibly damaging her lungs.' However, the question we have been trying to answer is whether the unacceptable doing could be attributed to the *relevant* agent. Therefore, insofar as we consider S (and not  $S^*$ ) to be the relevant agent, attribution of the doing to the relevant agent would, in this case, be unproblematic. This in turn would lead to the conclusion that the description of what is *going on* at  $t^*$  indeed refers to something unacceptable that is *being done by the relevant agent*. In the next section, I will get back to this conclusion and what it implies for Andreou's alternative view as a solution to the practical problem in intrapersonal trivial difference cases. Before doing so, however, let's consider what Andreou's

---

<sup>5</sup> There are many different views on the metaphysical relation between time slices and temporally extended agents (for a discussion, see Thoma 2018). Andreou is not explicit about which view she endorses. For the present purposes, I simply grant any view about time-slice agency that would allow for Andreou's view to do its work as a solution to the inefficacy paradox in the intrapersonal context. Ultimately, the claims I am making in this paper, regarding the (non-)applicability of this solution to the *interpersonal* context, do not depend on its usability in the *intrapersonal* context. Therefore, I take it that any objection to Andreou's alleged assumptions regarding time-slice agency, although it may threaten some of Andreou's arguments in the intrapersonal context, will not undermine my arguments here.

constraint of attributing extended doings only to extended agents implies for collective cases like *Herder*.

In *Herder*, we have seen that what is going on at  $t^*$ , when H grazes his herd, can be accurately described as an unacceptable doing: ‘irreversibly damaging the common pasture.’ Again, the question that needs to be answered is: does this description refer to something unacceptable that is being done *by the relevant agent*? In other words, can the unacceptable doing be attributed to the relevant agent? Just like in *Smoker*, we can distinguish between a temporally extended agent, H, who extends beyond  $t^*$ , and a time-slice agent,  $H^*$ , who persists only for as long as  $t^*$ . Following the same reasoning as above, it should be clear that ‘irreversibly damaging the common pasture,’ being a temporally extended doing, cannot be attributed to  $H^*$ , because it requires a temporally extended agent for its attribution. Since H is such a temporally extended agent (that is, we can assume he persists during the entire time the common pasture is being destroyed), does that mean that ‘irreversibly damaging the common pasture’ can be attributed to H? In what follows, I will argue that it cannot, due to the fact that this doing is not only *temporally* extended, but also *spatially* extended.

To understand this, it is worth looking in more detail at what precisely is needed to reach the conclusion that, in *Herder*, what is going on at  $t^*$  (when H grazes his oversized herd) can be described in terms of the unacceptable doing ‘irreversibly damaging the common pasture.’ Andreou (2014, 2022) appears to take for granted that if what is going on at  $t^*$  can be described as an *individual* doing in progress, then it can just as easily be described as a *collective* doing in progress. It should be noted, however, that the latter requires an additional step that has up to now been neglected. Merely recognizing the *temporal* extendedness of what is going on does not suffice, for this would allow us, at most, to describe what is going on at  $t^*$  as: ‘grazing a herd in the common pasture *many times*.’ This description still refers to an *individual* doing in progress, and not a *collective* one. And, by assumption, the damage that *one* (oversized) herd can do to the common pasture is trivial, so even if it were grazed *many times*, what is going on would not amount to ‘irreversibly damaging the common pasture.’ Therefore, it seems that describing what is going on at  $t^*$  in terms of the *collective* doing in progress ‘irreversibly damaging the common pasture’ requires that we recognize not only the temporal extendedness, but also the ‘*spatial*’ extendedness of what is going on. What exactly do I mean by spatial extendedness? If what is going on at  $t^*$  is

described as ‘grazing a herd in the common pasture *many times*,’ this description—while it indeed refers to what is going on over a longer period of time—only refers to what is going on *within* a certain, limited space. Namely, in this case, the space in which H happens to graze his herd.<sup>6</sup> Acknowledging the spatial extendedness of what is going on at  $t^*$ , then, means to recognize that what is going on is not limited to what can be completed *in* one specific space, but also includes things that happen *beyond* this space (that is, in the spaces where the other herders graze their herds). Only then can we describe what is going on in terms of the *collective* doing in progress: ‘grazing *many* herds in the common pasture *many times*,’ which, by description of the case, *does* amount to ‘irreversibly damaging the common pasture.’ So, in addition to the temporal extension, this spatial extension of the description is an essential step in concluding that what is going on at  $t^*$  is unacceptable (relative to the relevant concern of maintaining a viable common pasture).

Let’s now return to the question whether the unacceptable doing ‘irreversibly damaging the common pasture’ can be attributed to H. As I mentioned above, I want to argue that it cannot, given that the doing is not only temporally, but also spatially extended. But we still need to understand, then, how the spatial extendedness of a doing might affect its attributability. Even though the spatial extension has not been discussed by Andreou, I propose that we can nevertheless learn from her views on the attributability of temporally extended doings here. It should be noted that there is a structural analogy between the temporal extension and the spatial extension of the description of what is going on at  $t^*$ . In both cases, we are extending (albeit in a different dimension) this description either to include things that are going on beyond a particular *time point* or things that are going on beyond a particular *space*. Now, we have seen that, according to Andreou, the temporal extension came at the cost of being able to attribute the temporally extended doing to a non-temporally extended time-slice agent. Although not much additional argumentation is provided for this idea, it seems plausible that the reason why Andreou believes there to be this limitation for the attributability of temporally extended doings, is that such doings are constituted by actions (in this case, smoking many more cigarettes) executed at times when the time-slice agent does not ‘exist’ as such. In other words, most of the doing

---

<sup>6</sup> The term ‘space’ should not be understood here in the most literal sense. That is, it does not refer to a particular, static piece of the pasture, for it might well be that the exact location in which H’s herd grazes is different each time.

takes place beyond the ‘realm’ in which the time-slice agent can operate. If this is right, then, given the mentioned structural analogy between temporally and spatially extended doings, I propose that we might infer a similar limitation to be in place for the attribution of spatially extended, collective doings.

What kind of limitation precisely? In collective cases like *Herder*, there is an additional level of agency to be considered. That is, not only can we distinguish between the time-slice agent  $H^*$  and the temporally extended agent  $H$ , but also between the *individual* herder  $H$  and the *collection* of herders that he is a part of. Just like a time slice can be taken as a non-temporally extended agent, capable of operating only within a certain time point, we can understand an individual to be a ‘non-spatially extended’ agent, in the sense that he is able to operate only within a certain space.<sup>7</sup> Exactly like in the temporally extended case, then, we could therefore say that a collective doing—being, as we saw, spatially extended—takes place beyond the ‘realm’ in which an individual, non-spatially extended agent is able to operate. Therefore, I propose that, if I correctly interpreted Andreou’s underlying rationale for why a temporally extended doing cannot be attributed to a time-slice agent, it would have to be accepted that, similarly, a collective doing cannot be attributed to an individual.<sup>8,9</sup> Instead, when a doing in progress “reaches beyond” the individual’s space, this doing must be “attributable to an entity sufficiently [spatially] extended to carry out the doing” (Andreou 2022, 38). In other words: a *collection* of agents that, taken together, is able to operate in the entire space where the doing takes place.<sup>10</sup>

---

<sup>7</sup> This space, then, includes only those areas in which the agent can bring about (relevant) effects by virtue of what is under her control. In *Herder*, for example, the space in which  $H$  is able to bring about any relevant effects is limited by (among other things) the fact that he only controls one herd, which is unable to graze multiple patches of the pasture at the same time.

<sup>8</sup> Note that this also supports the (obvious) idea that, in the case of  $W$  crossing the street on his way to work, ‘ $D$  is driving her car’ is not a description of something that is being done by  $W$ . The description refers to a doing that is taking place outside the space in which  $W$  can operate, given that this space is limited by (among other things)  $W$ ’s inability to be present in two places at once.

<sup>9</sup> It seems unlikely that Andreou would disagree with my suggestion here, given that in the examples she provides to illustrate cases in which a *collective* doing is in progress, her descriptions of what is going on always involve reference to an entity that is in some way spatially extended. For example, in case two people are making a kite together, Andreou (2014) describes what is going on as “*M and I* are making a kite” (217, emphasis added), rather than describing *one* of the two individuals as ‘making a kite.’

<sup>10</sup> In light of the numerous debates surrounding collective agency (for an overview see Roth 2017; Schweikard and Schmidt 2021), I do not wish to make any claims about, for instance, the conditions under which a collection of individuals can be defined as a

So, if we now go back to the question that was originally posed, whether we can conclude that, since H is a temporally extended agent, the temporally extended doing ‘irreversibly damaging the common pasture’ can be attributed to H, the answer seems to be ‘no.’ Given its spatial extendedness, the doing ‘irreversibly damaging the common pasture’ is constituted by actions (namely, the grazing of many other herds in the common pasture) that are performed in spaces where the individual agent, H in this case, is unable to bring about any relevant effects. If I am right that we can infer from Andreou’s own constraint on attribution, that spatially extended, collective doings can only be attributed to spatially extended entities,<sup>11</sup> the doing ‘irreversibly damaging the common pasture’ could then only be attributed to the entire *collection* of herders, but not to H, the individual herder. Earlier we said that a description of what is going on only refers to something that is being done *by the relevant agent* if that doing can be attributed to that agent. At this point, we therefore must conclude that, insofar as we consider H to be the relevant agent, the description ‘irreversibly damaging the common pasture,’ albeit an accurate description of what is going on at  $t^*$ , does not refer to something unacceptable that is being done *by the relevant agent*. Call this the *attribution problem*.

---

collective agent, whether or not (intentional) collective agency is possible, and other matters along these lines. For the present purposes, what matters is only the (arguably) uncontroversial idea that a collection of agents can occupy and operate in a larger space than a single individual.

<sup>11</sup> Even if Andreou would disagree with my line of reasoning here, despite this being a natural way to conceive of the attribution of collective doings, I believe her own reasoning would lead her, via a different route, to draw the same conclusion. To see how, we have to look closer at *when* a particular description of a doing in progress would even qualify as an accurate answer to the question of what is going on at a time of interest. So far, I have not gone into this in much detail, but Andreou (2014) in fact does provide a more refined condition for when a certain description is appropriate: she argues that a particular doing, say X-ing, can be said to be in progress at  $t$ , only if it is reasonable to assume that this X-ing will “in due course” (215) and “barring interference” (215) result in having X-ed. For example, we can accurately describe W, at  $t^*$  when we see him crossing the street, as ‘walking to work’ only if it would be natural to assume that W will, at some point, have walked to work. Whether or not this *would* be a reasonable assumption, depends on W’s intentions, dispositions, and other relevant facts about the world. Now assume that the herders’ dispositions and the state of the common pasture are such that it would be appropriate to assume that they will, in due course, have irreversibly damaged the common pasture. We can then accurately describe what is going on, when an oversized herd is being grazed at  $t^*$ , as ‘*The herders* are irreversibly damaging the common pasture.’ Even so, given that by assumption, H’s individual use of the common pasture has trivial effects, it would *not* be reasonable to assume that, in due course, H will have irreversibly damaged the commons. Therefore, following Andreou’s own logic, it would not be accurate to describe what is going on at  $t^*$  as ‘H is irreversibly damaging the common pasture.’

## V. TAKING STOCK

Where does that leave us? We have been trying to answer the question of whether the alternative view of what is going on in trivial difference cases gives agents like S and H a reason to abstain from performing the trivial acts that, taken together, eventually lead to the outcome that conflicts with their relevant concern. If so, this would solve the practical conflict that arose in such cases under the standard view. Recall that we said that such a reason would be found if it could be argued that, were S to smoke another cigarette, or were H to graze his oversized herd in the common pasture, they would be acting unacceptably in doing so. We also said that an agent is considered to be acting unacceptably if an unacceptable doing (that is, a doing that would conflict with a relevant concern she has) can be attributed to her. So, we should see to what extent the aforementioned constraints on attributing doings to agents allow us to evaluate the acceptability of the agent's conduct based on the relevant descriptions of what is going on.

Starting with *Smoker*. The unacceptable doing in question here, which would be going on *at* each time S smokes a cigarette, is 'irreversibly damaging her lungs,' as it would frustrate S's concern with maintaining her lungs in good health. Recall that we established that this doing can be attributed to S, but not to the time-slice agent S\*. It should be noted that, in the present case, this observation is unproblematic: given that we are concerned with evaluating S's conduct at t\*, the question of what is being done by S\* is, for the current purposes, irrelevant. And, given that the unacceptable doing *can* be attributed to S at t\*, it seems that we can conclude that, when S smokes a cigarette, she would be acting unacceptably at that time, which gives S a reason not to smoke the cigarette.<sup>12</sup>

So far, so good. But can we draw the same conclusion for H? In *Herder*, the unacceptable doing would be 'irreversibly damaging the common pasture,' as this is the doing that conflicts with the relevant concern of maintaining the common pasture in good condition. Given the attribution problem, however, we saw that the unacceptable doing *cannot* be

---

<sup>12</sup> As was pointed out by an anonymous reviewer, it could be objected that even if there is now a reason for S *not* to smoke the cigarette, there is also still a reason for S in *favor* of doing so, given the fact that smoking a cigarette will bring her momentary pleasure. Not smoking the cigarette, it could be argued, would thus frustrate a concern she might have with obtaining this momentary pleasure. However, insofar as S values her good health over the pleasure that she would get from one cigarette (which I believe to be a plausible assumption), the reason *against* smoking it should be sufficiently strong for her not to do so.



attributed to H, but only to the *collection* of herders. This conclusion regarding the attribution of the unacceptable doing should make it clear that, even on the alternative view, we are not in a position to say that H would be acting unacceptably at the time he is grazing his oversized herd.<sup>13</sup> Therefore, it seems, the alternative view cannot, unlike in S's case, help provide a reason for H to refrain from grazing his oversized herd in the common pasture. Therefore, it must be concluded that the alternative view does not provide a satisfying solution to the practical problem in collective trivial difference cases.

Does this mean that the alternative view cannot help us *at all* in trivial difference cases in the collective context? Not exactly. Let me briefly say something about how Andreou's approach can provide valuable insights into a slightly different, though interrelated type of inefficacy paradox that cases like *Herder* have, on closer inspection, embedded within them. The problem that we have been addressing so far was that, given the trivial effects of a single instance of overgrazing by a single herder, it seems that this herder would not be doing anything unacceptable if he were to graze his herd in the common pasture. However, in cases like *Herder*, it can also plausibly be assumed that even if *all* the herders would engage in one instance of grazing their herds, the effect of what is *collectively* being done is still trivial (and therefore, what is collectively being done seems to be acceptable). It is only over time that their collective grazing habits gradually become problematic to the condition of the common pasture. So, even when considered as a unified collective, the herders together face the same problem H faced by himself: if they would not be doing anything unacceptable when all grazing their herds, say, today, what reason could they, as a collective of herders, have not to do so?<sup>14</sup> It should be easy to see how the alternative view could solve the predicament that the collective finds itself in: given that we can describe what is being done *at each time* they are all engaged in grazing their herds as 'irreversibly damaging the common pasture,' and given that, unlike H, the collective of herders *is* an entity that is sufficiently spatially extended to be attributed this collective doing, we can accurately say that the *collective of herders* would be acting unacceptably each time they are engaged in

---

<sup>13</sup> Some authors have attempted to argue that by *participating* in an unacceptable collective doing, an individual would automatically be acting unacceptably, despite the triviality of their individual contribution (for example see Parfit 1984; Kutz 2000). Such attempts, however, have yet to prove resistant against what I consider to be valid objections, such as those raised by Nefsky (2019).

<sup>14</sup> The problem presented here is described in more detail in Andreou (2006).

grazing their herds. The alternative view would therefore provide a reason *for the collective* not to overgraze their herds, and instead, for example, agree to change their grazing habits to be more sustainable.

Even so, H (and each of the other herders constituting the collective) still face a dilemma. Given that the collective doing cannot be attributed to him, and given that the effects of *his* grazing habits are assumed to make no perceptible difference to what is *collectively* being done (both *in t\** as well as *over time*), H still seems to find no reason why *he* should change his grazing habits or reduce the size of his herd. Given that the same would hold for each herder in the collective, it seems that the alternative view cannot solve this particular predicament.

## VI. CONCLUSION

In this paper, we have seen that Andreou's alternative characterization of trivial difference cases provides valuable insights into the inefficacy paradox that, on the standard view, arises in such cases. Contrary to the standard view, we no longer must conclude that, paradoxically, nothing unacceptable is going on at any point in time during which we inch our way toward a destructive outcome. We have also seen how, in *intrapersonal* cases of trivial difference, the alternative view implies that an agent thus finds a (previously lacking) reason to refrain from the individually trivial doings that, when repeated many times, lead to the destructive outcome. If Andreou's alternative view were accepted, it would thus constitute a promising solution to both the theoretical and the practical problem that arises in intrapersonal trivial difference cases. However, I have argued that this conclusion does not transfer to *collective* trivial difference cases, because the conditions for attributing doings to agents in such cases lead to the attribution problem. If my analysis is correct, then the application of the alternative view to collective trivial difference cases amounts, at best, to the conclusion that what we are doing *together* leads to unacceptable consequences, and that we should *collectively* change our conduct in such situations. This conclusion, however, seems somewhat redundant: it may well have been clear to each individual, from the outset, that what is being done *collectively* is problematic. Indeed, we often realize this full well in our everyday lives, when we are stuck in traffic or witness the results of environmental damage. The problem is that it is unclear why our *individual* behavior would count as problematic. Even after employing the alternative view, the practical problem that the inefficacy paradox gives rise to remains unsolved in the collective case, and

it therefore remains unclear what any individual should do differently. Ultimately, the effects of what is collectively being done can be reduced to the effects of what the individuals combined are doing, so it is precisely this problem that, in collective trivial difference cases, gives rise to the predicament that the agents face; and it is precisely this problem that Andreou's alternative view cannot seem to solve. I do not mean to imply that there exist no reasons *at all* for agents to avoid the trivial doings that jointly cause trouble. But, if my analysis here is correct, more arguments would be needed in order to conclude that such reasons are rooted in the notion that an individual performing such a doing would be frustrating the relevant concern in question. In sum, I have laid out a structural disanalogy between intrapersonal and collective trivial difference cases that limits the applicability of a solution to the inefficacy paradox from one type of case to the other. This insight may also prove useful for exploring the transferability of other attempted solutions to the inefficacy paradox from the intrapersonal to the collective context.

## REFERENCES

- Andreou, Chrisoula. 2006. "Environmental Damage and the Puzzle of the Self-Torturer." *Philosophy & Public Affairs* 34 (1): 95-108.
- Andreou, Chrisoula. 2014. "The good, the bad, and the trivial." *Philosophical Studies* 169 (2): 209-225.
- Andreou, Chrisoula. 2022. *Commitment and Resoluteness in Rational Choice*. Cambridge: Cambridge University Press.
- Anscombe, Elizabeth. 1963. *Intention*. Cambridge: Harvard University Press.
- Hardin, Garrett. 1968. "The Tragedy of the Commons." *Science* 162 (3859): 1243-1248.
- Korsgaard, Christine. 2008. *The Constitution of Agency*. Oxford: Oxford University Press.
- Kutz, Christopher. 2000. *Complicity: Ethics and Law for a Collective Age*. Cambridge: Cambridge University Press.
- Nefsky, Julia. 2019. "Collective harm and the inefficacy problem." *Philosophy Compass* 14 (4): 1-17.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Roth, Abraham Sesshu. 2017. "Shared Agency." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Article published December 13, 2010; last modified May 1, 2017. <https://plato.stanford.edu/archives/sum2017/entries/shared-agency>
- Schweikard, David P. and Hans Bernhard Schmid. 2021. "Collective Intentionality." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Article published June 13, 2013; last modified December 9, 2020. <https://plato.stanford.edu/archives/fall2021/entries/collective-intentionality>
- Sinnott-Armstrong, Walter. 2010. "It's Not My Fault: Global Warming and Individual Moral Obligations." *Climate Ethics: Essential Readings*, edited by Stephen M. Gardiner, Simon Caney, Dale Jamieson and Henry Shue, 177-190. Oxford: Oxford University Press.

- Soon, Valerie. 2021. "An intrapersonal, intertemporal solution to an interpersonal dilemma." *Philosophical Studies* 178 (10): 3353–3370.
- Tenenbaum, Sergio. 2023. "Rational Powers in Interaction: Replies to Paul, Andreou, Brunero, Mayr, and Haase." *Philosophical Inquiries* 11 (1): 163–183.
- Tenenbaum, Sergio and Raffman, Diana. 2012. "Vague Projects and the Puzzle of the Self-Torturer." *Ethics* 123 (1): 86–112.
- Thoma, Johanna. 2018. "Temptation and Preference-Based Instrumental Rationality." In *Self-Control, Decision Theory, and Rationality: New Essays*, edited by J.L. Bermúdez, 27–47. Cambridge: Cambridge University Press.
- Thompson, Michael. 2008. "Naïve action theory." In *Life and action: Elementary Structures of Practice and Practical Thought*, 27–47. Cambridge: Harvard University Press.
- Velleman, David. 2015. *The Possibility of Practical Reason*. Ann Arbor: Michigan Publishing.
- Watson, Gary. 1996. "Two Faces of Responsibility." *Philosophical Topics* 24 (2): 215–236.
- Wolf, Susan. 1990. *Freedom Within Reason*. Oxford: Oxford University Press.

**Tessa Supèr** is a PhD candidate at the Vrije Universiteit (VU) Amsterdam, working on rationality in temporally extended and collective action.  
Contact e-mail: <t.e.super@vu.nl>