Securing obligations: a reply to Hindriks

MATTIAS GUNNEMYR
University of Gothenburg

CAROLINE TORPE TOUBORG
Independent Researcher

Abstract: In his contribution to this special issue, Hindriks considers the Security Principle, an account of pro tanto obligations based on our account of reasons (Gunnemyr and Touborg 2023a). According to the Security Principle, you have a pro tanto obligation not to perform an action that makes a harm more secure. Hindriks raises two objections to this account. First, that it is too flexible; second, that it gives wrong verdicts when agents are robustly unwilling to act in a certain way. Here, we respond to Hindriks' objections and argue that Hindriks' account, the Threshold Principle, gives counterintuitive verdicts in preemption and low-probability cases.

Keywords: collective harm, pro tanto obligations, making a difference, security dependence, the threshold principle.

JEL Classification: A130, D70, D81

I. INTRODUCTION

In his contribution to this special issue, Hindriks considers the Security Principle:

\[(SP) \text{ There is a pro tanto obligation not to perform an action that makes a harm more secure.}\]

This principle is closely related to the account of reasons we present in “Reasons for action” (2023a). In the following, we first present our account of reasons and how it relates to \(SP\). Next, we consider Hindriks’ two
main criticisms of SP. Finally, we raise two worries for Hindriks’ account of when an action is impermissible for reasons of harm.

II. SECURITY, REASONS, AND PRO TANTO OBLIGATIONS

Think of some time when you barely made it in time for the bus, the doors closing just as you were entering. If there had been a red light at some crosswalk or if you had not hurried, you would have missed the bus. Here, the security of catching the bus was low: while this event—you catching the bus—did occur, it was very close to not occurring. Compare this to some time when you arrived at the bus stop well ahead of time. Then, it would not have mattered if you had encountered a red light or walked more slowly. You would have caught the bus anyway. In this case, you catching the bus was more secure: it was quite far from not occurring.

We spell out what it is for an event to be secure in terms of distances between possible worlds. When you barely made it to the bus, you catching the bus had a low degree of security: the distance between the actual world and the closest possible world(s) where you did not catch the bus was small. Conversely, when you arrived at the bus stop long before the bus, the nearest possible world(s) where you did not catch the bus was further away.

Just as we can assess the security (always positive) of an event that did occur, we can also assess the security (always negative) of an event that did not occur. Think of a time when you missed the bus by a hair—the doors closed just as you were running the final meters. Here, the event—you catching the bus—did not occur, but it was very close to occurring. At other times, you may have missed the bus by several minutes. Then, the event—you catching the bus—was very far from occurring.

According to our account of reasons, reasons are contrastive: you have a reason to \( \varphi \) rather than \( \psi \). Further, reasons are grounded in contrasts between two incompatible outcomes: a better outcome \( O \) (for example, catching the bus) versus a worse outcome \( O^* \) (for example, missing the bus). Roughly, you have a reason to \( \varphi \) rather than \( \psi \) just in case your \( \varphi \)-ing rather than \( \psi \)-ing makes the better outcome \( O \) more secure and the worse outcome \( O^* \) less secure. For example, when you want to catch the bus, you have a reason to leave your house early rather than late, because leaving early rather than late increases the security of catching the bus and decreases the security of missing the bus.

Following Nefsky (2017), we think that not all possibilities are relevant for determining what you have reason to do. In some distant possible
worlds, the Earth is attacked by aliens who are afraid of the color pink. Painting your house pink would make it more secure that you survive their attack. However, this obviously does not give you a reason to paint your house pink. The possibility of an attack by color-pink-fearing aliens is simply too remote to play any role in determining what you have reason to do. Or consider Nefsky’s example: even if there is a far-flung possibility that Obama will come by your house for tea this afternoon, this possibility is irrelevant for determining your reasons. Instead, we should only consider a restricted set of possible worlds—namely, those possible worlds that represent relevant possibilities. We call this set of possible worlds the relevant possibility horizon. More carefully, then, you have a reason to $\varphi$ rather than $\psi$ just in case your $\varphi$-ing rather than $\psi$-ing makes the better outcome $O$ more secure and the worse outcome $O^*$ less secure within the relevant possibility horizon.¹

Our account of reasons entails that you have a reason not to perform an action that makes a harm more secure (since the contrasts are obvious in the cases discussed here, we leave them out in what follows). The Security Principle (SP) makes a stronger claim, namely that you have a pro tanto obligation not to do so, where a pro tanto obligation (from here, ‘obligation’) is a “presumptively decisive reason” (Hindriks 2022, 837).

Consider driving an SUV on a Sunday afternoon. According to our account, you have a reason not to do so, because doing so increases the security of climate harm. But you may have other—and sometimes stronger—reasons that favor driving. For example, Hindriks considers a case where your wife is in labor and driving is the only way to get her to the hospital in time. Here, our account of reasons delivers the verdict that you have (at least) two reasons: a climate-harm-related reason not to drive and a (presumably stronger) getting-to-the-hospital-in-time-related reason to drive. Thus, our account does not commit us to the implausible claim that you have an obligation not to drive in this case. In most of the cases Hindriks considers, however, there are no such strong, countervailing reasons. Here, we are happy to accept SP.

Just as our account of reasons relies on the notion of a relevant possibility horizon, SP implicitly does so too. Fully spelled out, then, SP should be:

(SP) There is a pro tanto obligation not to perform an action that makes a harm more secure within the relevant possibility horizon.

¹ For more detail, see Gunnemyr and Touborg (2023a).
Thus, \( SP \) on its own does not offer a complete account of obligations. To get a full account, we need to combine \( SP \) with an account of the relevant possibility horizon.

### III. Hindriks’ First Objection: Flexibility

\( SP \) can be combined with many different accounts of what the relevant possibility horizon is. Depending on which account we choose, we may get very different verdicts. Consider the following case:

**Lake:** Alice, Bill, and Cecil live around a lake. Each has a boat. For several years, they have independently used cheap but toxic paint to paint their boats each spring. So far, the fish in the lake are thriving. However, the toxins are reaching a critical level. If two or more lakers paint their boat with cheap toxic paint this year, the fish will die. So, at least two must safely dispose of the old paint and buy a new, non-toxic paint to save the fish. However, each laker is robustly unwilling to stop using toxic paint. They are utterly indifferent to the ecosystem, and there is nothing anyone can do to change their minds. Finally, each laker knows all this.\(^2\)

In the following, we illustrate the verdicts of \( SP \) when it is combined with two different accounts of the relevant possibility horizon.

Suppose first that we combine \( SP \) with an account of the relevant possibility horizon that is not sensitive to robust unwillingness. Consider, for example, the following: it is a relevant possibility that you could do something just in case you would do it, if you chose to do it. Thus, it is, despite her robust unwillingness, a relevant possibility that Alice could abstain from using toxic paint, since she would do it, if she chose to do it. Similarly for Bill and Cecil. This generates the large possibility horizon below:

\[\text{[Large possibility horizon]}\]

\(^2\) As Hindriks notes, we consider a version of this case where the lakers are *not* robustly unwilling in Gunnemyr and Touborg (2023a) and (2023b). Björnsson (2014) originally introduced that version.
Given $H_{\text{Large}}$, $SP$ entails that each laker has an obligation not to use toxic paint. Consider Alice. The closest-to-$@$ world where Alice uses toxic paint is $@$; the closest-to-$@$ world where she abstains is $w_1$. While the fish die in both worlds, the bad outcome is more secure in $@$: to go from $@$ to a world where the fish do not die, two lakers would have to act differently; to go from $w_1$ to a world where the fish do not die, only one laker would have to act differently. By using toxic paint, Alice thus makes the harm (the death of the fish) more secure. Thus, Alice has an obligation not to use toxic paint (and similarly for Bill and Cecil).

However, we might instead combine $SP$ with a different account of the relevant possibility horizon, according to which the relevant possibility horizon for assessing Alice’s obligations is the small possibility horizon illustrated below:

Given $H_{\text{Alice}}$, $SP$ entails that Alice has no obligation not to use toxic paint (and mutatis mutandis, this applies to Bill and Cecil too): within $H_{\text{Alice}}$, the death of the fish is infinitely secure in both $@$ and $w_1$, since there simply is no world in $H_{\text{Alice}}$ where the fish do not die. Thus, Alice does not make the death of the fish more secure by using toxic paint, since it is already infinitely secure either way.
This illustrates how SP delivers very different verdicts depending on which account of the relevant possibility horizon we combine it with. Hindriks suggests this flexibility is “both a blessing and a curse” (this volume, aa).

Let us start with the blessing. The fact that the verdicts of SP depend on our choice of possibility horizon offers an attractive diagnosis of why intuitions vacillate in some cases—such as Lake—and why people may disagree about such cases. The idea is that vacillation or disagreement about obligations—for example, about whether Alice has an obligation not to use toxic paint—may be traced back to an underlying uncertainty or disagreement about which possibilities to include in the relevant possibility horizon. This diagnosis fits with the observation that emphasizing certain features of the situation may sway intuitions. For example, emphasizing the fact that there is a pattern of behavior that allows the fish to survive—namely, when two or more lakers use non-toxic paint—makes $H_{\text{large}}$ more salient and prompts the intuition that Alice (and the others) has an obligation not to use toxic paint. By contrast, emphasizing that, given the robust unwillingness of Bill and Cecil, the fish will die no matter what Alice does makes $H_{\text{Alice}}$ more salient—and this prompts the intuition that Alice has no obligation.

Now to the curse: Hindriks objects that SP is too flexible. He suggests that, if we simply leave it up to people’s intuitions which possibilities to include in the relevant possibility horizon, SP “will confirm whatever intuitions you have” (this volume, aa). We agree that this would not be an attractive feature.

In response, the key is to note that SP is incomplete. To get a complete account of obligations, we need to combine SP with a principle for deciding which possibilities to include in the relevant possibility horizon. In Lake, for example, this might tell us that the relevant possibility horizon is $H_{\text{large}}$ and, correspondingly, deliver the unequivocal verdict that Alice does have an obligation not to use toxic paint. Such an account still allows us to offer a psychological explanation of why you may think Alice has no obligation—namely, if you take $H_{\text{Alice}}$ to be the relevant possibility horizon. However, thinking this would simply be a mistake.

In “Reasons for action,” we have already made some initial suggestions about which possibilities to include in the relevant possibility horizon. We will now attempt to expand on this.

We take our starting point in Nefsky’s (2017) suggestion that “the possibilities that come into the account are those that we should regard as
live possibilities in practical deliberative contexts” (2760). When a single person deliberates about what to do, she is typically trying to choose between several different courses of action. These are the courses of action she takes to be her options. Suppose, for example, that Alice is deliberating about what to have for dinner. Here, she may take her options to be: (a) going shopping, or (b) ordering take-out from the local restaurant.

In some cases, of course, an agent may be mistaken about her options: she may fail to realize that a course of action she takes to be an option is, in fact, closed off, or she may fail to notice that some further course of action is also an option for her. Alice may fail to realize the local restaurant is closed for the holidays, so ordering take-out is not an option, or she may fail to notice she already has enough food in the fridge for a third option to be available—namely, (c) cooking something with what she has.³

However, even after such mistakes are corrected, a deliberating agent will often, correctly, take herself to have several options to choose between. Roughly, something is an option for an agent just in case she would do it, if she chose to. We say ‘roughly’ because this may need to be finessed in cases involving compulsions and phobias, and possibly also in other cases.

In a situation involving many agents, each will typically (correctly) take herself to have several options to choose between, and we will think about them as having such options. As Nefsky writes: “we think of agents (both ourselves and others) as typically being able to choose between several different courses of action, where different outcomes can result depending on what they choose to do” (2762). Whenever a course of action is an option for an agent, we think the corresponding possibility should be included in the relevant possibility horizon. This yields the following rough account (Gunnemyr and Touborg 2023a, 344):

Relevant possibilities: The possibility of an agent acting in a particular way should be included in the relevant possibility horizon just in case it is an option for her to act in that way.

We believe SP should be completed by combining it with Relevant possibilities. However, this leads to Hindriks’ second objection.

---

³ Jeppsson (2016) discusses such cases.
IV. HINDRIKS’ SECOND OBJECTION: ROBUST UNWILLINGNESS

When we combine SP with Relevant possibilities, we get the verdict that Alice has an obligation not to use toxic paint in Lake: Even though she is robustly unwilling to use non-toxic paint, doing so is presumably an option for her. She is not prevented from doing so by any phobia, addiction, or compulsion, and if she were to attempt to use non-toxic paint, she would succeed. The same holds for the other lakers. According to Relevant possibilities, then, the relevant possibility horizon should include the possibility of each of the lakers using non-toxic paint. Thus, $H_{\text{large}}$ is the relevant possibility horizon and, as seen above, this delivers the verdict that Alice (and each of the others) has an obligation not to use toxic paint.

If the lakers were not robustly unwilling, Hindriks would presumably be happy to accept that Alice (and each of the others) has an obligation not to use toxic paint. According to Hindriks, however, the robust unwillingness of the lakers means it simply is not feasible to save the fish, and because of that, he thinks Alice has no obligation in this version of the case:

None of them is obligated to change their behavior if all of them are robustly unwilling. To be sure, had there even been a remote possibility that the lakers would have changed their mind, then each might have been obligated to stop using the solvent. However, in the situation at issue, saving the fish is not a real option. (Hindriks, this volume, aa)

Therefore, he objects that SP gives counterintuitive verdicts in cases like Lake.

This objection does not threaten SP itself: as we have seen above, SP can deliver different verdicts, depending on which principle for deciding the relevant possibility horizon it is combined with. Consider, for example, Relevant possibilities*:

Relevant possibilities*: To determine an agent’s pro tanto obligations, the possibility of other agents acting differently should not be included in the relevant possibility horizon when they are robustly unwilling to do so.

---

4 Björnsson (2014) agrees with Hindriks' intuition.
Combining $SP$ with *Relevant possibilities* delivers the verdict Hindriks prefers: according to *Relevant possibilities*, the relevant possibility horizon for determining Alice’s obligations is $H_{Alice}$, and as seen above, this entails that Alice has no obligation, since there is no possibility of saving the fish. Even if you agree with Hindriks’ verdict about *Lake*, then, you do not need to reject $SP$. Instead, you may simply question which principle for deciding the relevant possibility horizon it should be combined with and, in particular, how this principle should treat (robust) unwillingness.

Can unwillingness to do something relieve you of an obligation to do so? In the individual case, this seems false. A child should do her homework even if she does not want to, and you should attend the meeting you promised to participate in even if you are unwilling to do so. Unwillingness subtracts nothing from our obligations in these cases. As Lawford-Smith (2013) writes: “we don’t want to let agents off the moral hook. The fact that a person won’t do something isn’t enough for us to retract an imperative that she ought to” (254). This applies in collective cases too: when we consider all three lakers at once, it seems clear that they ought to save the fish, even though they are unwilling to do so.

The intuition that Alice has no obligation only appears when we consider what she should do, given the robust unwillingness of the others. Alice might think: “I could use toxic paint or abstain from doing so. But given Bill and Cecil’s robust unwillingness, there is no possibility the fish will survive. So, it just doesn’t matter what I do”. This perspective corresponds to $H_{Alice}$, where we hold fixed what Bill and Cecil will do, while treating it as a relevant possibility that Alice might act in different ways. From this perspective, Alice has no obligation. Similarly, if we take Bill’s perspective—holding fixed what Alice and Cecil will do, while treating it as a relevant possibility that Bill might act in different ways—we find that Bill has no obligation. The same applies to Cecil.

However, there is a reason to reject these perspectives. The reason is that these three perspectives are at odds with each other. From Alice’s perspective, we treat it as a relevant possibility that she might use non-toxic paint, but not that Bill (or Cecil) might do so. From Bill’s perspective, by contrast, we treat it as a relevant possibility that he might use non-toxic paint, but not that Alice (or Cecil) might do so. According to one

---

5 Björnsson (2014) agrees that the three lakers together have an obligation to save the fish, even if they are unwilling to do so. However, he suggests that while they have a shared obligation, none of them has an individual obligation.

6 For a discussion on how taking different perspectives can influence intuitions, see Jeppsson (2016).
perspective, then, it is a relevant possibility that Alice could use non-toxic paint; according to the other, it is not. In this way, the perspectives contradict each other: we cannot accept both (or all three) of them at once.

This reveals that, while Alice’s perspective initially seems compelling, there is something problematic about it. Alice’s perspective is biased in a certain way: although the situation of the three lakers is perfectly symmetric, Alice treats herself differently from the others—she treats it as a relevant possibility that she could abstain from using toxic paint, but not that Bill (or Cecil) could. But if Alice sees it as a relevant possibility that she could use non-toxic paint, and Bill sees it as a relevant possibility that he could use non-toxic paint, they should also recognize these facts about each other (and so should we). This supports Relevant possibilities and points towards the larger possibility horizon, \( H_{\text{Large}} \).

Perhaps, though, robust unwillingness goes deeper. So far, we have imagined that Alice sees herself as facing a choice between using toxic or non-toxic paint: while she is unwilling to use non-toxic paint, she still sees it as a relevant possibility that she could do so. But maybe that does not fully capture what Hindriks has in mind. Maybe Alice is robustly unwilling in the sense that she does not even see it as an option to use non-toxic paint—in her view, doing so is not a relevant possibility. Similarly for Bill and Cecil. In this version, there is no tension between the perspectives of the three lakers: all three agree on a possibility horizon that contains just a single world, namely @ where the fish die.

In this case, we could still claim that Alice, Bill, and Cecil are simply mistaken: while they do not take themselves to have the option of using non-

---

7 In Gunnemyr and Touborg (2023a), we suggest that the Principle of Moral Harmony lends independent support to the verdict that each laker has an obligation not to use toxic paint. Hindriks suggests that Lake presents a counterexample to the principle. However, this objection (mistakenly) presupposes that the right verdict about Lake is that the lakers lack obligations.
toxic paint, in fact they do. Because of this, it is a relevant possibility that each of them uses this option, and thus, the relevant possibility horizon remains $H_{\text{large}}$.

However, there is an alternative approach. As long as you are a member of the community of moral agents, it makes sense to treat you as someone who can have obligations and, therefore, to treat it as a relevant possibility that you could care about what matters and act accordingly. But sometimes, someone may be so far beyond the reach of moral considerations that we should no longer treat her as a member of the community of moral agents—at least in the situation at hand. When someone is excluded from the community of moral agents in this way, there is not even a question about what her obligations are (in that situation)—just as there is no question about whether a blackbird has an obligation to sing at night. Correspondingly, reasoning with her or starting a blame conversation does not make sense.

Alice might be like this. She might be completely insensitive to considerations about the lake’s ecosystem. If we were to challenge her about using toxic paint, she would not even attempt to defend herself—she would not think she had to. Instead, she might say something like, “Seriously?! I couldn’t care less about those silly fish. There’s just no way I could be bothered to use a different paint”. If so, we should not treat her as a member of the community of moral agents (at least not in this situation), and the question about her obligations does not even arise. In such a case, $SP$ does not apply.

When we exclude someone from the community of moral agents, we are, in some sense, giving up on them—we do not even hold on to the possibility that they could care and act accordingly. In some cases, however, we need to cut our losses: give up on one or more individuals to focus on what can still be achieved by those who remain within the community of moral agents.

Excluding one or more individuals from the community of moral agents can make a difference to the obligations of the agents who remain within the community. Consider a variation of Lake: Bill and Cecil are robustly unwilling to use non-toxic paint. Their unwillingness is so deep and

---

8 An exception to this would be if they are unwilling in a way that amounts to something like an obsessive-compulsive disorder. Then it is not an option for them to use non-toxic paint. Here, the appropriate possibility horizon is $H_{\text{small}}$, and we agree they have no obligation.

9 We take the community of moral agents to include, roughly, all individuals who can respond to moral reasons.
intractable that we should exclude them from the community of moral agents. But in this version, Alice remains within the community. What are her obligations in this case? Here, \( H_{\text{Alice}} \) is indeed the relevant possibility horizon—and as we already know, \( SP \) then entails that Alice has no obligation to use non-toxic paint.

V. OBJECTIONS TO HINDRIKS’ ACCOUNT

We now turn to Hindriks’ preferred account, the Threshold Principle:

\[
(\text{TP}) \quad \text{There is a pro tanto obligation not to perform an action exactly if the probability that it brings about a harm is too high.}
\]

More specifically, Hindriks instructs us to consider two conditional probabilities: (1) the probability that the harm will occur, given that you perform the action (while everyone else acts as they actually do), and (2) the probability that the harm will not occur, given that you do not perform the action (while everyone else acts as they actually do). If both probabilities are high enough (that is, above a threshold), you have an obligation not to perform the action.

The virtues of \( TP \) are clear in cases where a harm depends probabilistically on the actions of several people, as in Hindriks’ case of a dance club, where the probability that someone will faint gradually increases as the dance club gets more and more packed. In preemption cases, however, \( TP \) delivers counterintuitive verdicts. Consider the following case:

**Sandcastle**: A child is building an elaborate sandcastle on the beach. Suzy throws a ball at it, and so does Billy. Suzy’s ball hits the sandcastle first. When Billy’s ball hits, the sandcastle is already destroyed, and the child is distraught. If Suzy had not thrown her ball, Billy’s would have destroyed the sandcastle and the child would still have been distraught.

Suppose that if Suzy throws her ball, it is certain (probability equal to 1) that the sandcastle will be destroyed and the child distraught, and similarly for Billy.

Suzy should not have thrown her ball at the sandcastle. She had no good reason for doing so; refraining would not have been costly for her, and so on. However, this is not what \( TP \) entails. The probability that the child will be distraught if Suzy throws is 1. This is, presumably, above the threshold. However, the probability that the child will not be distraught if
Suzy does *not* throw is 0, because of Billy’s throw. This is, presumably, below the threshold. According to TP, you only have an obligation to refrain from an action when both probabilities pass their respective thresholds. Thus, TP entails that Suzy has no obligation not to throw her ball at the sandcastle. This spells trouble for TP, and the same difficulties arise in overdetermination cases too. By contrast, SP easily delivers the desired verdict in these kinds of cases: in Sandcastle, Suzy’s throw makes the harm more secure; therefore, she has an obligation not to throw.\(^{10}\)

TP also faces difficulties in low-probability cases. Consider the following case:

*Lousy Thrower Suzy*: This time, only Suzy is around, and she is a miserable thrower, or perhaps the sandcastle is just far away. Anyway, the probability that she will hit the sandcastle if she throws is very low. She knows this and throws her ball. Luckily for her, but unfortunately for the child, she hits and destroys the sandcastle, and the child is distraught.

Suzy should not have thrown her ball at the sandcastle. That she was unlikely to hit it is no excuse. However, the probability that harm will occur given that Suzy throws is very low. Indeed, given any threshold higher than zero, we can simply tweak the case to ensure that the probability of harm is just below the threshold. If so, TP counterintuitively implies that Suzy had no obligation not to throw her ball. By contrast, SP can deliver the desired verdict: Suzy’s throw makes the harm more secure, since the harm actually occurs and would not have occurred if she had not thrown.

VI. CONCLUSION

Hindriks makes two objections to SP. His first objection is that SP is too flexible—it “will confirm whatever intuitions you have” (this volume, aa). In response, we have argued that, to get a complete account of pro tanto obligations, SP needs to be combined with a principle for deciding which possibilities to include in the relevant possibility horizon. Hindriks’ second objection concerns the verdicts we get when we combine SP with our preferred account of relevant possibilities. According to Hindriks, this yields counterintuitive verdicts in cases where people are robustly

---

\(^{10}\) You might think Suzy has no obligation, given that Billy threw his ball. We can explain this intuition as the result of (mistakenly) adopting a smaller possibility horizon, where Billy’s throw is held fixed.
unwilling to act in a certain way. We consider two responses to this objection. One response is to emphasize that unwillingness does not relieve you from obligations in the individual case, and it seems that the same applies in collective cases. If so, SP delivers correct verdicts in such cases, *pace* Hindriks. However, there is also an intriguing second response: it may sometimes be appropriate to exclude one or more individuals from the community of moral agents (in a particular situation). When someone is excluded from the community of moral agents, there is not even a *question* about what her obligations are (in that situation), but the exclusion of some individuals from the community of moral agents may change the obligations of those who remain.

**REFERENCES**


**Mattias Gunnemyr** is a postdoctoral researcher in the Financial Ethics Research Group at the Department of Philosophy, Linguistics and Theory of Science at the University of Gothenburg. He specializes in the ethics of inefficacy, particularly focusing on climate ethics, financial ethics and causation.

Contact e-mail: <mattias.gunnemyr@gu.se>

**Caroline Torpe Touborg** is an independent researcher working on causation, responsibility, and climate change. She has previously been a postdoctoral researcher with the Lund Gothenburg Responsibility Project and at Umeå University.

Contact e-mail: <caroline.touborg@gmail.com>