



Reasons for action: making a difference to the security of outcomes

Mattias Gunnemyr¹ · Caroline Torpe Touborg²

Accepted: 12 August 2022
© The Author(s) 2022

Abstract

In this paper, we present a new account of teleological reasons, i.e. reasons to perform a particular action because of the outcomes it promotes. Our account gives the desired verdict in a number of difficult cases, including cases of overdetermination and non-threshold cases like Parfit's famous *Drops of water*. The key to our account is to look more closely at the metaphysics of causation. According to Touborg (*The dual nature of causation*, 2018), it is a necessary condition for causation that a cause increases the security of its effect. Building on this idea, we suggest, roughly, that you have a teleological reason to act in a certain way when doing so increases the security of some good outcome. This represents a middle way between the proposal that you have a reason to act in a certain way just in case this *would* cause a good outcome and the proposal that you have a reason to act in a certain way just in case this *could* cause a good outcome.

1 Introduction

Many of the problems society faces today involve multiple agents: each individual agent makes no perceptible difference, but the result of thousands or millions of people acting in a particular way may nevertheless be catastrophic or save us all. Climate change is perhaps the most obvious example, but there are many more. The following case, originally presented by Parfit (1984), brings out the crucial features and works as a metaphor for many other such cases:

✉ Caroline Torpe Touborg
caroline.touborg@umu.se; caroline.touborg@gmail.com

Mattias Gunnemyr
mattias.gunnemyr@fil.lu.se

¹ Department of Philosophy, Lund University, Lund, Sweden

² Department of Historical, Philosophical and Religious Studies, Umeå University, Umeå, Sweden

Drops of water: ‘Imagine that there are ten thousand men in the desert, suffering from intensely painful thirst. We are a group of ten thousand people near the desert, and each of us has a pint of water. We can’t go into the desert ourselves, but what we can do is pour our pints into a water cart. The cart will be driven into the desert, and any water in it will be evenly distributed amongst the men. If we pour in our pints, the men’s suffering will be relieved. The problem is, though, that while together these acts would do a lot of good, it does not seem that any individual such act will make a difference. If one pours in one’s pint, this will only enable each man to drink an extra *ten thousandth* of a pint of water. This is no more than a single drop, and a single drop more or less is too minuscule an amount to make any difference to how they feel.’ (Nefsky, 2017: 2743–44).

We seem to lack a clear intuitive verdict about what an individual has reason to do in cases like this. On the one hand, there is a strong intuition that each of us has a reason to donate our pint of water. After all, the men’s suffering will be alleviated if enough of us donate our pints. On the other hand, donating a single pint will not make any difference to the men’s suffering – so how can you have a reason to do so?

We side with the intuition that each of us has a reason to donate our pint (our motivation for this verdict is given in Section 2.). However, it has proved difficult to find a general account of reasons for action that supports this intuition. One of the most promising proposals is presented by Nefsky (2017). As Fanciullo (2020) shows, however, Nefsky’s account faces serious counterexamples. In this paper, we therefore present a new account of reasons for action that supports the desired verdict on *Drops of water*.

More precisely, our account supports the claim that each of us has an *objective* and *pro tanto* reason to donate our pint. We focus on *objective* reasons for action in the sense that we defend the claim that each of us has a reason to donate our pint, even when all relevant information is taken into account. And we focus on *pro tanto* reasons in the sense that we defend the claim that we each have a reason to donate our pint, that is, there is a consideration that speaks in favour of doing so.¹ Whenever we merely talk of reasons in the following, this is shorthand for objective *pro tanto* reasons.

The claim that we each have a *pro tanto* reason to donate our pint is weaker than the claim that we each have an obligation to do so. Still, our account of reasons may provide a key input to the question whether we each have an obligation to donate our pint. This is so because the most obvious challenge to the claim that we each have an obligation to donate our pint is that donating an extra pint makes no difference, so we have no *reason* to do so. As Nefsky writes:

¹ If donating your pint has costs, you also have a *pro tanto* reason not to donate your pint. In the following, we set aside questions about how to determine the strength of reasons, how one should weigh up *pro tanto* reasons in order to arrive at *all-things-considered* reasons, and whether we can arrive at obligations by weighing up reasons.

‘When one says, ‘but it won’t make *any difference*’, more than just saying, ‘it doesn’t seem that I am obligated to act in that way’, one is saying ‘there doesn’t seem to be *any point at all* in acting in that way.’ (Nefsky, 2017: 2744–45).

By showing that we each have a reason to donate our pint, we respond to the challenge: we show that there *is* a point in adding an extra pint.

To show this, we develop a unified account of *teleological* reasons for action.² A teleological reason to ϕ is a reason to ϕ that is grounded in the fact that ϕ -ing *promotes* a certain outcome – for example, the outcome that the men’s suffering is relieved (see e.g. Portmore 2018a).³ A crucial question in understanding when you have a teleological reason to ϕ is: what is the relevant relation of *promoting*?

In the following, we propose a new answer to this question. Roughly, we propose that *promoting* should be understood in terms of *making an outcome more secure* (see Sect.5). Even if the men’s suffering is not in fact fully relieved, donating your pint makes this good outcome more secure: it brings us a step closer to achieving the good outcome. This proposal is motivated by three considerations. First, it successfully handles cases that present problems for rival accounts. Second, it is theoretically motivated: it establishes a clear connection between causing and promoting, and it upholds key inferences. Third, it allows us to understand why intuitions sometimes vacillate.

We proceed as follows. First, we motivate why we side with the verdict that each of us has a reason to donate our pint in *Drops of water* (Sect.2). We then set out three starting assumptions about reasons for action (Sect.3), and we argue that an account of promoting needs to capture certain key inferences and deliver the correct verdicts in a number of test cases (Sect.4). Next, we set out our account and show that it upholds the key inferences (Sect.5) and delivers the desired verdicts in our test cases (Sect.6). We end by showing that our account delivers the desired verdict in *Drops of water* (Sect.7) and that, in addition, it can explain why some cases elicit vacillating intuitions (Sect.8).

2 Motivating our verdicts: the principle of moral harmony

In the following, we shall consider a number of cases where intuitions vacillate. We have already seen such vacillation in the case of *Drops of water*: on the one hand, there is a strong intuition that you have a reason to donate your pint; on the other

² We leave it open whether *all* reasons are teleological or whether there are also non-teleological reasons. An example of a non-teleological reason might be a reason based on a promise: the fact that you have promised to ϕ may provide a reason for you to ϕ , which is independent of whether ϕ -ing promotes some good outcome. Since we leave room for non-teleological reasons, our account is compatible with a wide range of first-order moral theories. Note that whenever we speak of reasons in the following, these are teleological reasons.

³ Standardly, reasons are taken to be facts. For example, the fact that it is raining is a reason for you to bring an umbrella. This way of thinking about reasons may be connected to questions about promoting as follows: a fact F is a reason for you to ϕ just in case F explains why your ϕ -ing promotes some good outcome. For example, the fact that it is raining explains why bringing an umbrella promotes the outcome that you remain dry. See e.g. Schroeder (2007).

hand, it seems that donating your pint makes no difference – so how could you have a reason?

When intuitions vacillate in this way, we cannot simply require that our account should respect intuitions. Instead, we need some independent support for the verdicts we side with. We may get such support from the principle of moral harmony.⁴ Applied to objective reasons, the principle states that if all the agents involved in a situation act as they have objective reason to act, then the resulting pattern of behaviour will lead to the best attainable outcome. By contraposition, we find that if a suboptimal outcome is produced, there must be at least one agent who has *failed* to act as she had objective reason to act (see Pinkert 2015: 975–77).

We may think of this as a principle about explanation: when a suboptimal outcome occurs in a situation where a different pattern of behaviour would have produced an optimal outcome, the principle tells us that there is an explanation in terms of one or more agents failing to act as they had objective reason to act. It will be useful to refer back to this principle, and we therefore set it out here:

Explaining suboptimal outcomes: If a suboptimal outcome occurs in a situation where a different pattern of behaviour would have produced an optimal outcome, then at least one of the agents involved in the situation has failed to act as she had objective reason to act.

It seems reasonable to assume that a corresponding principle holds in the case of optimal outcomes:

Explaining optimal outcomes: If an optimal outcome occurs in a situation where a different pattern of behaviour would have produced a suboptimal outcome, then at least one of the agents involved in the situation has acted as she had objective reason to act.

These principles provide the guidance we need by imposing constraints on the verdicts that an account of reasons should deliver: when a suboptimal outcome occurs (and a different pattern of behaviour would have produced an optimal outcome), our account of reasons should deliver the verdict that at least one of the agents involved had an objective reason to do something other than what she in fact did. When an optimal outcome occurs (and a different pattern of behaviour would have produced a suboptimal outcome), our account of reasons should deliver the verdict that at least one of the agents involved had an objective reason to act as she in fact did.

⁴ Proponents of the principle include Regan (1980), Parfit (1984: 54), Pinkert (2015), Portmore (2018b), and Fanciullo (2020: 1488–89). The exact formulation of the principle varies between different authors. To be fully precise, we endorse the following version of the principle: for any world w , if all the agents involved in a situation act as they have objective reason to act in w , then the resulting pattern of behaviour will lead to the best attainable outcome in w . In his criticism of the principle, Feldman (1980) seems to have an alternative version in mind, namely: let w be the closest possible world where all the agents involved in a situation act as they have objective reason to act in @; then the resulting pattern of behaviour will lead to the best attainable outcome in w . We agree with Feldman that this version of the principle is untenable (see footnote 21).

The two principles support the verdict that you have a reason to donate your pint in *Drops of water*. Suppose first that no one donates their pint. In that case, a sub-optimal outcome occurs (the men's suffering continues unmitigated) in a situation where a different pattern of behaviour (e.g. everyone's donating their pint) would have produced an optimal outcome (the full alleviation of the men's suffering). Here, *Explaining suboptimal outcomes* tells us that at least one of the agents involved in the situation has failed to act as she had objective reason to act: at least one of the agents has failed to donate her pint when she had an objective reason to do so. Since there are no relevant differences between the agents, we should conclude that each agent had an objective reason to donate her pint. Correspondingly, if we suppose that everyone donates their pint and the optimal outcome is achieved (the men's suffering is fully relieved), *Explaining optimal outcomes* supports the verdict that at least one agent (and, by symmetry, every agent) had an objective reason to donate her pint: the optimal outcome was achieved *because* everyone acted in accordance with their reasons. This gives us the guidance we need when choosing whether to trust the intuition that you have a reason to donate your pint or the intuition that you do not.

We now turn to the challenge of developing a general account of reasons for action that can support the verdict that you have a reason to donate your pint in *Drops of water*. To do so, we will need to set aside *Drops of water* for a while: in order to get a general account of reasons, we need to consider a wide range of cases. Before turning to other cases, however, we first set out our starting assumptions.

3 Starting assumptions

For the sake of simplicity, we assume in the following that the laws of nature are deterministic. Furthermore, we rely on three assumptions about teleological reasons.

First, we assume that the actions you have reason to do are time-indexed (see e.g. Skorupski 2010): you do not simply have a reason to ϕ ; rather, you have a reason to ϕ at time t . Or, in the case of temporally extended actions or omissions, you may have a reason to *begin* to ϕ at time t .⁵

Second, we assume, following Snedegar (2017), that teleological reasons are contrastive: you do not simply have a reason to ϕ at t ; you have a reason to ϕ rather than ψ at t , where ϕ and ψ are two incompatible actions or omissions in the sense that it is not possible for you to both ϕ and ψ at t . Furthermore, you do not merely have such a reason in virtue of how your action relates to some outcome O ; you have such a reason in virtue of how your action relates to whether outcome O will occur *rather than some incompatible outcome* O^* . Our assumption that teleological reasons are contrastive can be motivated by the very same cases that motivate a contrastive account of causation, such as:

Train tracks: Suppose that you are standing by a switch in the railroad tracks. The switch has three settings: *express*, *local*, and *broken*. If the switch is set to

⁵ There is a further question about *when* you have a reason to ϕ at t : do you only have this reason at t , or do you also have it some time prior to t , or even after t ? We stay neutral on this question.

express, the train will arrive quickly at the station; if the switch is set to *local*, the train will arrive slowly at the station; and if the switch is set to *broken*, the train will derail. Suppose that of these three outcomes, the best outcome is that the train arrives quickly at the station; the second-best is that the train arrives slowly at the station; and the worst outcome is that the train derails. Suppose further that the switch is initially set to *broken*. (Cf. Schaffer 2012: 38)

Some intuitions about this case may be difficult to capture without the resources of contrastivism. Consider the following two claims:

- (i) You have a reason to move the switch to *local* rather than moving it to *express*.
- (ii) You have a reason to move the switch to *local* rather than leaving it at *broken*.

It seems clear that (i) is false while (ii) is true. Since the only difference between these two claims lies in the choice of contrast, we need to go contrastive in order to accommodate both of these verdicts.⁶ To avoid cumbersome repetitions, however, we will sometimes leave out the contrasts in what follows, saying simply that ‘you have a reason to ϕ ’ when it is obvious what the relevant contrast ψ is.

Third, we assume that you only have a reason to ϕ rather than ψ at time t when it is an option for you to ϕ at t and an option for you to ψ at t . We understand the notion of an option in a natural, commonsense way, where it is often true that you have several different options open to you at a given time. In *Drops of water*, for example, you have at least two options: you have the option of donating your pint and the option of keeping it to yourself. However, it is not an option for you to, for instance, go back in time and singlehandedly prevent slavery, the crusades, and the two world wars. Since doing so is not an option for you, you do not have a reason to do so, no matter how much good could be achieved if you succeeded. In this way we avoid what Streumer (2007) calls ‘crazy reasons.’

Given these three assumptions, we may state our question more precisely. The question is how to fill in the blank in the schema below in a way that captures how your ϕ -ing rather than ψ -ing promotes the occurrence of outcome O rather than O^* .⁷

SCHEMA: You have a teleological reason to ϕ rather than ψ at time t , where ϕ and ψ are two mutually incompatible actions or omissions, just in case

- (a) it is an option for you to ϕ at t ,
 - (b) it is an option for you to ψ at t , and
- there are two incompatible outcomes O and O^* , such that
- (c) O is better than O^* , and
 - (d) [fill in the blank].

⁶ Sinnott-Armstrong (2008) argues in favour of contrastivism about reasons in a similar fashion. For further discussion, see Snedegar (2017: 25–44).

⁷ Our aim is to provide a general account of reasons for action, including moral, prudential, and aesthetic reasons. To get an account specifically of *moral* reasons, we need to place a restriction on condition (c) so that it requires that O is better than O^* in a *morally significant way*.

4 Desiderata: key inferences and test cases

In this section, we consider three prominent suggestions about how to fill in the blank in SCHEMA. By understanding the advantages and disadvantages of these suggestions, we can get a better picture of the desiderata that a successful account of teleological reasons needs to satisfy. The three suggestions are:

Whether-whether dependence:

(d) whether O or O* occurs depends on whether you ϕ or ψ at time t .

Cause:

(d) your ϕ -ing rather than ψ -ing at time t would be a cause of O rather than O*.

Potential cause:

(d) your ϕ -ing rather than ψ -ing at time t could be a cause of O rather than O*.

Consequentialists are typically committed to a non-contrastive version of *Whether-whether dependence*; Braham and van Hees (2012) defend a non-contrastive version of *Cause*; and Nefsky (2017) defends a non-contrastive version of *Potential cause*.⁸

Let us begin by considering *Whether-whether dependence*. We think this suggestion successfully captures a *sufficient* condition for when you have a reason. Correspondingly, the following inference holds: if it is the case that O would occur if you were to ϕ at t and O* would occur if you were to ψ at t , you have a reason to ϕ rather than ψ at t . Holding on to this inference is a desideratum for any account of teleological reasons.

However, *Whether-whether dependence* fails to give a necessary condition for when you have a reason. This becomes clear already in simple overdetermination cases like the following:

Nuclear safety: You and Suzy work as engineers at a nuclear power plant. You independently notice that there is a problem. At time t you each press a button to safely shut down the reactor. Each button-pressing is an overdetermining cause of the shut-down of the reactor. If just one of you had pressed your button at time t , the reactor would still have shut down safely. But if neither of you had pressed your button at time t , there would have been a nuclear disaster.

Our intuitions about *Nuclear safety* vacillate. On the one hand, it might seem that each of you had a reason to press your safety-button: if you had both failed to do so, there would have been a nuclear disaster. On the other hand, one might argue that since Suzy in fact pressed *her* button, *you* had no reason to press yours: given that

⁸ A further, natural suggestion is that promoting should be understood in terms of probability-raising (Schroeder, 2007). In a deterministic setting, this suggestion is extensionally equivalent to *Whether-whether dependence*.

Suzy pressed her button, the nuclear disaster would have been averted whether you pressed your button or not.

The principle of *Explaining optimal outcomes* supports the first intuition, namely that you each had a reason to press your safety-button: *Nuclear safety* describes a situation where an optimal outcome occurs (the reactor is shut down safely) and where a different pattern of behaviour would have produced a suboptimal outcome (a nuclear disaster). It therefore follows from the principle that at least one of you acted as you had objective reason to act. Since you and Suzy are symmetrically placed, the two of you must have the same reasons. Thus, the principle requires us to say that each of you had an objective reason to press your safety-button.

However, *Whether-whether dependence* delivers the verdict that *neither* you nor Suzy had a reason to press your buttons. Since Suzy pressed *her* button, it did not depend on your actions whether the reactor would be shut down safely or there would be a nuclear disaster. And since you pressed *your* button, it did not depend on Suzy's actions either.⁹

One alternative is to endorse *Cause* instead of *Whether-whether dependence* (as e.g. Braham and van Hees do). Again, we think that *Cause* succeeds in capturing a *sufficient* condition for when you have a reason. Correspondingly, the following inference holds: if it is the case that your ϕ -ing rather than ψ -ing at time t would be a cause of O rather than O*, you have a reason to ϕ rather than ψ at time t . Holding on to this inference is once again a desideratum for an account of teleological reasons.¹⁰

Cause satisfies *Explaining optimal outcomes* in *Nuclear safety*, provided that it is combined with an account of causation that allows for overdetermining causes: if we count your button-pressing as a cause of the safe shutdown, it immediately follows from *Cause* that you had a reason to press your button. Similarly, Suzy had a reason to press her button. However, we do not have to look far to find cases that create trouble for *Cause* as well. Consider, for example, the case below:

The lake: You, Vanessa and Walter all live close to a lake with a sensitive ecosystem. Each of you have a boat. If two or more of you paint the hull of your boat with a cheap and toxic paint rather than a non-toxic but more expensive one, the ecosystem in the lake will collapse. If at most one of you uses the toxic paint, the ecosystem will continue to thrive. As it turns out, all three of you use the cheaper paint, and the lake becomes a wet wasteland. (Adapted from Björnsson 2011 and 2014)

⁹ Consequentialists like Singer (1980), Norcross (2005) and Kagan (2011) are in principle committed to *Whether-whether dependence*. Thus, they are committed to saying that you did not have an *objective* reason to press your button in *Nuclear safety*. They mitigate the cost of this position by pointing out that you did have a *subjective* reason to press your button: assuming that you did not know what Suzy would do, the expected utility of pressing your button was higher than the expected utility of not pressing your button, since pressing your button *might* make the difference between safely shutting down the reactor and a nuclear disaster.

¹⁰ Nefsky (2017) uses a counterexample (the parking meter example, p. 2752) to argue that *Cause* does not give a sufficient condition for having a reason. In arriving at her verdict on what causes what in the counterexample, Nefsky relies on causal transitivity. However, as many have pointed out, transitivity fails precisely in cases with this structure (see e.g. Paul & Hall 2013: 232–244). We therefore believe her argument fails.

Did you, as an individual, have a reason to use the non-toxic paint rather than the toxic one? Once again, intuitions vacillate. On the one hand, it seems attractive to say that you had a reason to use the non-toxic paint rather than the toxic one and that each of the other two boat owners had such a reason too. On the other hand, one might note that since the other two boat owners in fact used the toxic paint, the ecosystem would have collapsed no matter what you did.

At this point, we turn to the principle of *Explaining suboptimal outcomes* for guidance: *The lake* describes a situation where a suboptimal outcome occurs (the ecosystem collapses) and where an alternative pattern of actions (at least two of you using the non-toxic paint) would have produced an optimal outcome (the ecosystem's continuing to thrive). Thus, the principle tells us that at least one of you has failed to act as you had objective reason to act: this failure explains the suboptimal outcome. Since the three of you are symmetrically placed, we should say that each of you had an objective reason to use the non-toxic paint rather than the toxic one.¹¹

However, *Cause* yields the verdict that you did not have such a reason. Remember that *Cause* says, roughly, that you have a reason to use the non-toxic paint just in case your doing so *would* be a cause of the ecosystem's continuing to thrive. If you had used the non-toxic paint in *The lake*, however, Vanessa and Walter would still have used the toxic paint, and the ecosystem would still have collapsed. Your using the non-toxic paint therefore would not have been a cause of the ecosystem's continuing to thrive for the simple reason that the ecosystem would not have continued to thrive. Thus, *Cause* yields the verdict that you had no reason to use the non-toxic paint in *The lake*. The same applies to Vanessa and Walter. *Cause* therefore implies that none of you had a reason to use the non-toxic paint rather than the toxic one.

Finally, let us consider *Potential cause*, which is the least demanding of the three conditions. *Potential cause* is satisfied in both of the cases we have considered so far: since your pressing your button *was* a cause of the reactor shutting down safely, rather than melting in a nuclear disaster, it obviously *could* be. And likewise, your using the non-toxic paint *could* be a cause of the ecosystem's thriving rather than collapsing, since it would be a cause if Walter or Vanessa had also used the non-toxic paint. *Potential cause* thus satisfies *Explaining optimal outcomes* and *Explaining suboptimal outcomes* in the cases we have considered, and indeed it does so generally.

However, whereas *Whether-whether dependence* and *Cause* encounter problems because they are too demanding, *Potential cause* runs into trouble because it is not demanding enough. To illustrate this, consider the following case:

¹¹ We may compare this verdict with the verdicts of Hindriks' account (2022). Hindriks' account is concerned with *pro tanto* obligations rather than reasons. However, we may ask: what would an account of reasons based on the central ideas of Hindriks' account look like? Hindriks considers the *prospect* of success, or, more carefully, the prospect of successfully helping to bring about a good outcome. This suggests the following rough account of reasons: you have a reason to ϕ just in case the prospect of your successfully helping to bring about the good outcome by ϕ -ing is good enough. A variation of *The lake* shows that this Hindriks-inspired account of reasons does not satisfy *Explaining suboptimal outcomes*: suppose that all three boat owners are unwilling to use the non-toxic paint. In that case, Hindriks would say that the prospect of success is not good enough (see Hindriks 2022: 438–41): given the unwillingness of the others, it is futile to try to save the ecosystem. This delivers the verdict that none of the three boat owners has a reason to use the non-toxic paint, a verdict that is at odds with *Explaining suboptimal outcomes*.

Coordination: As part of a game-show, you and Sally are placed on opposite sides of a wall. Each of you is given a choice between either raising your hand or not at a given signal. If you both raise your hands, or both do not, you will receive a million dollars each. If one of you raises your hand, and the other does not, you will receive nothing. You raise your hand at the given signal, and Sally does too. You win a million dollars each.

In this case, it is clear that you had a reason to raise your hand when the signal was given (call this time t) rather than keeping it down. It is also attractive to hold that you did *not* have any reason to keep your hand down at time t . Indeed, if you had kept your hand down at time t , you and Sally would both have received nothing.

However, *Potential cause* yields the verdict that you had both a reason to raise your hand and a reason to keep your hand down at time t : presumably, it is a relevant possibility that Sally might have kept her hand down at time t . And if Sally had kept her hand down, your keeping your hand down at time t would have caused each of you to receive a million dollars rather than nothing. Thus, keeping your hand down *could* have been a cause of the good outcome. In this case, we think that *Potential cause* admits reasons that simply are not there.

Nefsky's (2017) account is a version of *Potential cause*. According to Nefsky, you have a reason to ϕ just in case ϕ -ing *could* help, where helping consists in making a non-superfluous causal contribution. Not surprisingly, *Coordination* therefore presents a problem for Nefsky's account: in the possible world where neither you nor Sally raise your hands, keeping your hand down makes a non-superfluous causal contribution to your getting a million dollars each. Thus, keeping your hand down *could* help. And so, Nefsky's account yields the counterintuitive result that you have an objective reason to keep your hand down at t .

We now have clear desiderata for an account of teleological reasons. First, it needs to respect the sufficiency of *Whether-whether dependence* and *Cause*. Second, it needs to accommodate the desired verdicts (supported by *Explaining optimal outcomes* and *Explaining suboptimal outcomes*) in our three test cases: *Nuclear safety*, *The lake*, and *Coordination*. In order to achieve this, we need to find a condition that represents a middle way between *Cause* and *Potential cause* by being less demanding than *Cause* but more demanding than *Potential cause*. We set out how to do this in the following section.

5 Finding a middle way between *Cause* and *Potential cause*

To find a middle way between *Cause* and *Potential cause*, we pay closer attention to the metaphysics of causation: *Potential cause* is a weaker version of *Cause* because it merely requires that your ϕ -ing rather than ψ -ing at t *could* (rather than *would*) be a cause of O rather than O^* . However, once we pay attention to the metaphysics of causation, another attractive way of weakening *Cause* comes into view: if there are several necessary and jointly sufficient conditions for causation, we may weaken *Cause* by requiring only that some, but not all, of these conditions are satisfied. That is precisely the idea we develop in the following.

We begin from the account of causation developed by Touborg (2018). According to this account, there are two individually necessary and jointly sufficient conditions for causation: first, the condition of *process-connection*, which requires that a cause must be connected to its effect via a genuine process; second, the condition of *security-dependence within a possibility horizon*, which requires that a cause must make a difference to the security of its effect. We may weaken *Cause* by merely requiring that one of these two conditions is satisfied, rather than requiring the satisfaction of both. The condition of security-dependence within a possibility horizon is a highly promising candidate for this move.¹²

Consider again *Nuclear safety* where you and Suzy both press your safety buttons and where one such pressing is sufficient for safely shutting down the reactor. Even though neither you nor Suzy made a difference as to *whether* the outcome was going to occur, there is a sense in which each of you made the safe shutdown of the reactor *more secure*. As it happened, two things stood in the way of a nuclear disaster: your pressing your button and Suzy's pressing hers. If you had not pressed your button, only one thing would have stood in the way of a nuclear disaster: Suzy's pressing her button. The same reasoning applies to Suzy. In this way, each of you increased the security of the safe shutdown of the reactor. We think it is precisely because of this that both of you had a reason to press your safety buttons: by pressing your button rather than not, you made the safe shutdown of the reactor *more secure* while making a nuclear meltdown *less secure*.¹³

To make this more precise, it is useful to think about security in terms of the distance-at-a-time between possible worlds. Considering two possible worlds w and w^* , let us say that w is close-to- w^* -at-time- t to the extent that the complete state of world w at t is similar to the complete state of world w^* at t . Based on this, we may give an initial definition of security as follows:¹⁴

Security (initial definition): The security of outcome O in w at t is given as follows:

If O occurs in w , then O has *positive* security in w , and its degree of positive security in w at t is given by the distance-at- t between w and the closest-to- w -at- t world(s) where O does not occur.

If O does not occur in w , then O has *negative* security in w , and its degree of negative security in w at t is given by the distance-at- t between w and the closest-to- w -at- t world(s) where O occurs.

¹² While the idea for this move springs from the metaphysics of causation, our account of reasons is in principle independent of how the debate about causation turns out.

¹³ Pettit (2015) argues that the central ethical notions of attachment, virtue, and respect have an important modal component: what distinguishes e.g. a real friend from a fair-weather friend is a certain kind of modal robustness – if conditions had been more difficult, the real friend would still be there for you. If Pettit is correct in suggesting that such central ethical notions should be understood in terms of modal properties, this may make it easier to accept our suggestion that reasons should be understood in terms of the modal property of security. We are grateful to an anonymous reviewer for pointing out the similarity between our account and Pettit's work.

¹⁴ If you worry that we cannot refer determinately to an event O that did not occur, speaking of *types* of events solves this difficulty. More carefully, we should therefore say 'If an event of *type* O does not occur in w , then' For simplicity, we suppress this complication in the text.

With this initial definition, our proposal is going to be, roughly, the following way of filling in the blank in SCHEMA, using '@' to denote the actual world:

Security-dependence:

(d) O is more secure and O* is less secure at t in the closest-to-@-at- t world(s) where you ϕ at t than they are in the closest-to-@-at- t world(s) where you ψ at t .

To make this proposal fully precise, we need to answer the following question: which worlds should be taken into consideration when we are looking for 'the closest-at- t -worlds where ...'? Should we consider *all* possible worlds, in the widest sense, or should we make use of a restricted notion of possibility?

We suggest with Nefsky (2017) that only some possibilities are relevant for determining what you have reason to do. We may use the notion of a *possibility horizon* to capture this: a *possibility horizon* is simply a class of possible worlds, and the *relevant possibility horizon* for the purpose of determining what you have reason to do at time t is the class of worlds that contains just those possible worlds that are relevant for determining what you have reason to do at t . We suggest the following procedure for arriving at the relevant possibility horizon: Consider the agents involved in the situation in question. Each of these agents has certain actions and omissions that are open to her at time t . These are the agent's *options* at time t . A choice assignment is a specification of how each agent chooses among the options that are open to her at t .¹⁵ In the case of *Nuclear safety*, for example, there are four choice assignments, representing every combination of your choice {press, do not press} and Suzy's choice {press, do not press}. We believe that every such choice assignment represents a relevant possibility (see Nefsky 2017: 2762, fn. 37). As a minimum, the *relevant possibility horizon* $H(t)$ for determining what you have reason to do at time t should include worlds representing every such choice assignment.¹⁶ By contrast, we should not treat it as a relevant possibility that someone might do something that is not an option for her. Concerning non-agential features of the situation, we take it as a rule of thumb that it is not a relevant possibility that such features might be different from what they are like in the actual world. However, this is merely a rule of thumb: although we do not consider any such cases here, there may well be cases in which it is relevant to consider further possibilities involving alternatives to non-agential features of the situation.¹⁷

¹⁵ In some cases, some combinations of options may be impossible. For instance, it is not an option for you to dance the tango with me if I'm not willing to dance the tango with you. In this paper, we do not consider options that involve joint action. This means that, in the cases we discuss, every combination of the options we consider is possible.

¹⁶ Of course, merely specifying e.g. what you and Suzy do – for example, that neither of you press your safety button – is not enough to fully characterize a possible world. We assume that the relevant world(s) representing this possibility start out by being as similar as possible (consistent with neither of you pressing your button) to the actual world at time t , and then evolve forwards from there in accordance with the actual laws of nature. See the method for evaluating counterfactuals proposed in Paul & Hall (2013: 47–49).

¹⁷ For example, Fanciullo (2020) considers a version of *Drops of water* where the 9,999 other agents are replaced by mechanisms. He reports that 'my intuition that you can help in 9999 Mechanisms is, admit-

We can now complete our definition of security by relativising it to a possibility horizon:

Security within a possibility horizon: The security of outcome O in w at t within possibility horizon $H(t)$ is given as follows:

If O occurs in w , then O has *positive* security in w , and its degree of positive security in w at t within $H(t)$ is given by the distance-at- t between w and the closest-to- w -at- t world(s) in $H(t)$ where O does not occur.

If O does not occur in w , then O has *negative* security in w , and its degree of negative security in w at t within $H(t)$ is given by the distance-at- t between w and the closest-to- w -at- t world(s) in $H(t)$ where O occurs.

Based on this, we suggest the following completion of SCHEMA:

REASON: You have a teleological reason to ϕ rather than ψ at time t , where ϕ and ψ are two mutually incompatible actions or omissions, just in case

- (a) it is an option for you to ϕ at t ,
- (b) it is an option for you to ψ at t , and

there are two incompatible outcomes O and O*, such that

- (c) O is better than O*, and
- (d) O is *more secure* and O* is *less secure* at t in $H(t)$ in the closest-to-@-at- t world(s) in $H(t)$ where you ϕ at t than they are in the closest-to-@-at- t world(s) in $H(t)$ where you ψ at t , where $H(t)$ is the *relevant possibility horizon* for the purpose of determining what you have reason to do at t .

Assuming that you are in a situation where condition (a), (b), and (c) of REASON are satisfied, REASON entails that the following inferences hold (see [Appendix](#)):

The whether-whether inference:

If *whether* O or O* will occur depends on *whether* you ϕ or ψ at t , then you have a reason to ϕ rather than ψ at t .

The causal inference:

If your ϕ -ing rather than ψ -ing at time t *would* be a cause of O rather than O*, then you have a reason to ϕ rather than ψ at t .

REASON thus supports the key inferences we identified in Sect.4.

tedly, somewhat weaker than my intuition that you can help in Drops of Water' (p. 1493). Our account can capture this. In Fanciullo's 9999 Mechanisms, our rule of thumb does not settle what the relevant possibility horizon is. If the relevant possibility horizon only includes the options that are open to *you*, our account delivers the verdict that you have no reason to add your water. However, the relevant possibility horizon may also include further possibilities that are not based on the options open to agents: it may include the possibility that each mechanism contributes a pint or fails to do so. With this larger possibility horizon, our account delivers the verdict that you have a reason to donate your pint. Nefsky, by contrast, gets the verdict that you have no reason either way.

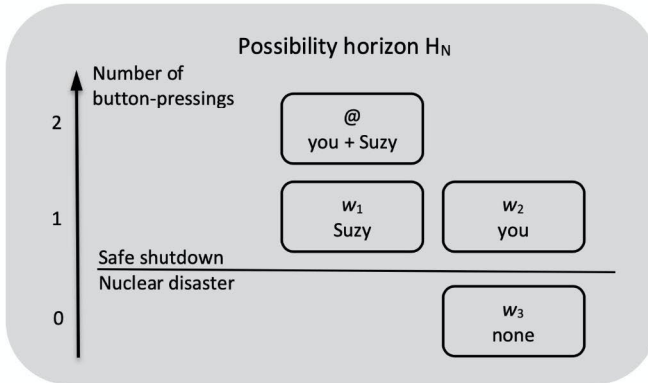


Fig. 1 The figure depicts the relevant possibility horizon for determining what you have reason to do in *Nuclear safety*, namely possibility horizon H_N with four possible worlds: $@$ where you and Suzy both press your buttons; w_1 where only Suzy presses her button; w_2 where only you press your button; and w_3 where neither you nor Suzy presses your button. The reactor is shut down safely in $@$, w_1 , and w_2 ; there is a nuclear disaster in w_3

6 Testing the account

In this section, we show that REASON delivers the desired verdicts in our three test cases – *Nuclear safety*, *The lake*, and *Coordination* – as well as in *Train tracks*.

First, consider *Nuclear safety*. Clearly, the first two conditions of REASON are satisfied: (a) it was an option for you to press the safety button at time t , and (b) it was an option for you not to press the safety button at time t . Furthermore, (c) is satisfied for the following choice of O and O^* :

- O = safe shutdown of the reactor
- O^* = nuclear disaster

From here, we could simply appeal to *The causal inference*: since your pressing your button rather than not is a cause of the reactor's shutting down safely, *The causal inference* delivers the result that you had a reason to press your button. However, we may also show directly that condition (d) is satisfied:

To do so, we first identify the relevant possibility horizon. You and Suzy each have two options: pressing your safety button or not. This means that the relevant possibility horizon as a minimum includes $2^2 = 4$ possible worlds, as illustrated in Fig. 1.

To see that (d) is satisfied, we need to consider two worlds: the closest-to- $@$ -at- t world within H_N where you press your button, namely $@$, and the closest-to- $@$ -at- t world within H_N where you do not press your button, namely w_1 .

Is the safe shutdown of the reactor *more secure* in $@$ than in w_1 ? Yes. The safe shutdown of the reactor occurs, and thus has positive security, in both $@$ and w_1 . However, there is a difference in its degree of security. Relative to $@$, the closest-at- t world where the reactor is not shut down safely is world w_3 . Relative to w_1 , the

closest-at- t world where the reactor is not shut down safely is still w_3 . Clearly, the distance-at- t between @ and w_3 is greater than the distance-at- t between w_1 and w_3 : @ and w_1 both differ from w_3 in terms of whether or not Suzy presses her safety button; in addition, @ also differs from w_3 in terms of whether or not you press your safety button. Thus, the safe shutdown is more secure in @ than in w_1 .

Is the nuclear disaster *less secure* in @ than in w_1 ? Yes. The nuclear disaster does not occur, and therefore has negative security, in both @ and w_1 . Once again, however, there is a difference in degree. The closest-to-@-at- t world where there is a nuclear disaster is world w_3 , and the closest-to- w_1 -at- t world where there is a nuclear disaster is also world w_3 . As we have seen, the distance-at- t between @ and w_3 is greater than the distance-at- t between w_1 and w_3 . This means that the nuclear disaster has a higher degree of negative security in @ than in w_1 : the nuclear disaster is, so to speak, further from happening in @ than in w_1 . Therefore, the nuclear disaster is *less secure* in @ than it is in w_1 , just as it is *less warm* when the temperature is -20° than it is when it is -10° .

This shows that condition (d) is satisfied. Thus, you have a reason to press your button in *Nuclear safety*. A parallel argument shows that Suzy has such a reason too.¹⁸

Let us next consider *The lake*. Do you have a reason to use the non-toxic paint rather than the toxic paint at the time t when you, Vanessa, and Walter are painting your boats?

Clearly, condition (a) and (b) are satisfied: (a) it is an option for you to use the non-toxic paint at t , and (b) it is an option for you to use the toxic paint. Furthermore, condition (c) is satisfied when

O = survival of the ecosystem

O* = collapse of the ecosystem

since the survival of the ecosystem is better than its collapse.

To show that (d) is satisfied, we first need to identify the relevant possibility horizon. There are three agents, each with two options: using the non-toxic paint at t or using the toxic paint at t . Thus, our possibility horizon should as a minimum include every combination of these courses of action, i.e. $2^3 = 8$ possible worlds, as illustrated in Fig. 2.

Within this possibility horizon, the closest-to-@-at- t world where you use the non-toxic paint at t is world w_1 , and the closest-to-@-at- t world where you use the toxic paint is @. The survival of the ecosystem has negative security in both w_1 and @. The closest-to- w_1 -at- t world(s) where the ecosystem survives are w_4 and w_5 , where you and one other boat-owner use the non-toxic paint. The closest-to-@-at- t world(s) where the ecosystem survives are w_4 , w_5 , and w_6 , where two people use the non-toxic paint. The distance-at- t between w_1 and w_4 or w_5 is *smaller* than the distance-at- t between @ and w_4 , w_5 , or w_6 : to get from w_1 to w_4 or w_5 , only one person needs to change which paint they are using, but to get from @ to w_4 , w_5 , or w_6 , two people need to change. This means that the survival of the ecosystem is *more secure* in w_1

¹⁸ Note that REASON delivers the same verdict in variations of *Nuclear safety* where only one of you or none of you press the safety-button.

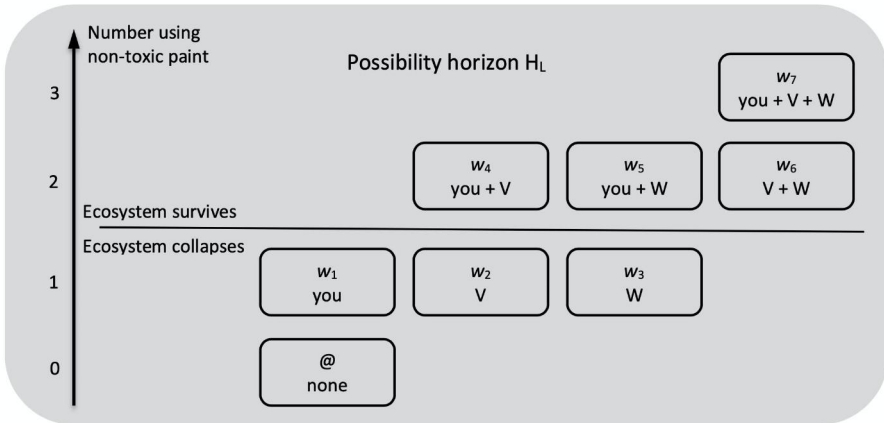


Fig. 2 The figure depicts the relevant possibility horizon for determining what you have reason to do in *The lake*, namely possibility horizon H_L with eight possible worlds: @ where none of the boat-owners use the non-toxic paint; w_1 , w_2 , and w_3 where one of the boat-owners uses the non-toxic paint; w_4 , w_5 , and w_6 where two of the boat-owners use the non-toxic paint; and w_7 where all three use the non-toxic paint. The ecosystem collapses in @, w_1 , w_2 , and w_3 ; the ecosystem survives in w_4 , w_5 , w_6 , and w_7

than it is in @: even though the ecosystem does not survive in either w_1 or @, it is *closer* to surviving in w_1 . A parallel argument shows that the collapse of the ecosystem is *less secure* in w_1 , where you use the non-toxic paint, than it is in @, where you use the toxic one. Thus condition (d) is satisfied, and we find, as we should, that you have a reason to use the non-toxic paint.¹⁹

Our third test case is *Coordination*. Here, we need to verify that REASON delivers the two verdicts we want: first, that you had a reason to raise your hand; and second, that you did *not* have a reason to keep your hand down. To show this, we need to consider the relevant possibility horizon, H_C , illustrated in Fig. 3.

We first show that you had a reason to raise your hand rather than keep it down: conditions (a) and (b) are clearly satisfied. Setting

O = each of you getting a million dollars

O* = getting nothing

condition (c) is also satisfied. By *The whether-whether inference*, we then find that you had a reason to raise your hand: O occurs in the closest-to-@-at- t world where you raise your hand (namely @), while O* occurs in the closest-to-@-at- t world (namely w_1) where you keep your hand down.

Importantly, REASON also delivers the verdict that you did *not* have a reason to keep your hand down rather than raising it. For in this case, condition (d) is not satisfied: it is not the case that your getting a million dollars each is *more secure* in the

¹⁹ REASON delivers the same verdict in variations of *The lake* where more boat owners use the non-toxic paint.

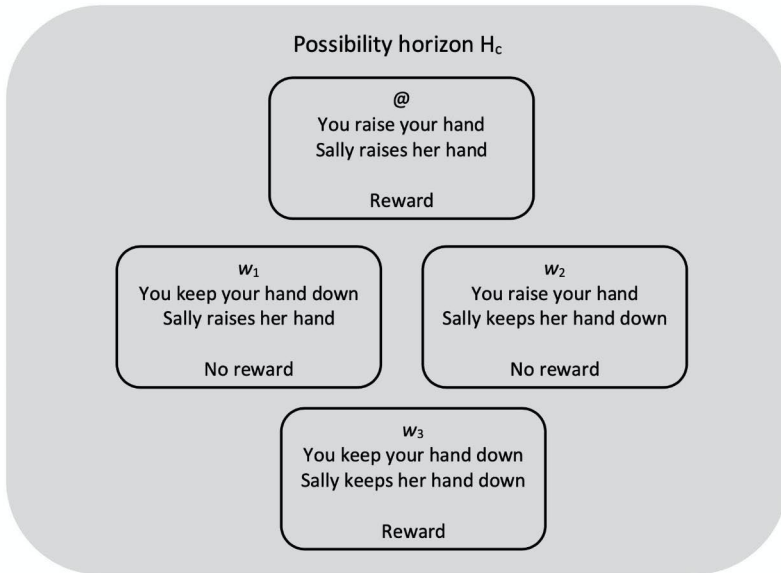


Fig. 3 The figure depicts the relevant possibility horizon for determining what you have reason to do in *Coordination*, namely possibility horizon H_c with four possible worlds: @ where you and Sally both raise your hands; w_1 where you keep your hand down while Sally raises hers; w_2 where you raise your hand while Sally keeps hers down; and w_3 where you both keep your hands down. You get a million dollars each in @ and w_3 ; you get nothing in w_1 and w_2

closest-to-@-at- t world where you keep your hand down, i.e. in w_1 , than it is in the closest-to-@-at- t world where you raise your hand, i.e. in @. On the contrary, your getting a million dollars each does not occur in w_1 (that is, it has negative security in w_1), while it does occur in @ (that is, it has positive security in @). Thus, REASON avoids the problem we identified for *Potential cause*. It does so by being sensitive to what in fact happened in the actual world – namely, that Sally raised her hand.²⁰

In the cases we have considered so far, we found that whenever O was *more secure*, O^* was *less secure*. Indeed, this is so whenever either O or O^* occurs in *every world*

²⁰ We noted earlier (footnote 4) that there are different versions of the principle of moral harmony. Our version goes as follows: for any world w , if all the agents involved in a situation act as they have objective reason to act in w , then the resulting pattern of behaviour will lead to the best attainable outcome in w . This version is straightforwardly satisfied in every variation of *Coordination*, i.e. independently of which world in the possibility horizon is the actual world. To compare, the version Feldman considers (and rejects) goes as follows: let w be the closest possible world where all the agents involved in a situation act as they have objective reason to act in @; then the resulting pattern of behaviour will lead to the best attainable outcome in w . It is easy to find a variation of *Coordination* where Feldman's version fails to be satisfied: suppose that @ is such that you keep your hand down while Sally raises hers. In that case, it seems intuitively clear that you had an objective reason in @ to raise your hand and Sally had an objective reason in @ to keep her hand down (and these are also the verdicts that REASON supports). In the closest world w where you both act as you have objective reason to act in @, Sally thus keeps her hand down while you raise yours. Obviously this leads to a suboptimal outcome in w . Thus, the version of the principle that Feldman considers fails to be satisfied. We agree with Feldman that this version of the principle should be rejected. We are grateful to an anonymous reviewer for prompting us to clarify this.

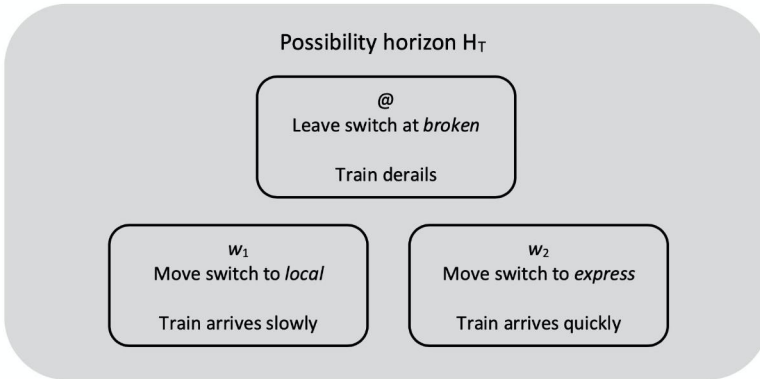


Fig. 4 The figure depicts the relevant possibility horizon for determining what you have reason to do in *Train tracks*, namely possibility horizon H_T with three possible worlds: @ where you leave the switch at *broken* and the train derailed; w_1 where you move the switch to *local* and the train arrives slowly; and w_2 where you move the switch to *express* and the train arrives quickly

within $H(t)$ (see the proof of *Symmetry* in [Appendix](#)). This might make you think that the requirement that O^* should be *less secure* in the closest ϕ -ing worlds than in the closest ψ -ing worlds is superfluous. However, this requirement does essential work in delivering the correct verdict in *Train tracks*:

The challenge in *Train tracks* is to simultaneously accommodate the falsity of (i) and the truth of (ii):

- (i) You have a reason to move the switch to *local* rather than moving it to *express*.
- (ii) You have a reason to move the switch to *local* rather than leaving it at *broken*.

To see how REASON accommodates these verdicts, the first step is to identify the relevant possibility horizon. This is illustrated in [Fig. 4](#). (We arbitrarily suppose that you leave the switch at *broken*, denoting the world in which you do so '@.' REASON delivers the same verdicts on the supposition that you move the switch to *local* or *express*, so that w_1 or w_2 becomes the actual world. This follows from *Stability*, which we prove in [Appendix](#).)

In the case of both (i) and (ii), we find that there is only one choice of O and O^* , such that condition (c) and the part of (d) that is concerned with O is satisfied:

- O = slow arrival
- O^* = derailling

Here, (c) is satisfied, since the train's slow arrival is better than its derailling. Furthermore, the train's slow arrival is *more secure* in w_1 (where you move the switch to

local) than it is in either @ or w_2 (where you leave the switch at *broken* or move it to *express*). The action lies in the part of (d) that is concerned with O^* :

In the case of (i), the relevant comparison is between the closest-to-@-at- t world where you set the switch to *local*, namely w_1 , and the closest-to-@-at- t world where you set the switch to *express*, namely w_2 . And we find that the train's derailing is *just as secure* in w_1 as it is in w_2 . Thus, the part of condition (d) that is concerned with O^* fails to be satisfied, yielding the intuitively correct verdict that (i) is false.

In the case of (ii), there is a different contrast – namely, leaving the switch at *broken*. This means that we have to make a different comparison: the relevant comparison is between w_1 and the closest-to-@-at- t world where you leave the switch at *broken*, namely @. Here, we find that the train's derailing is *less secure* in w_1 than it is in @. Thus, condition (d) is fully satisfied, and we get the intuitively correct verdict that (ii) is true.

This shows that REASON satisfies our desiderata: it delivers the desired verdicts in all our test cases, and, as we have seen above (Sect.5), it entails that *Whether-whenever dependence* and *Cause* give sufficient conditions for having teleological reasons. In the following section, we finally show how REASON delivers the desired verdict in *Drops of water*.

7 Drops of water

In *The lake*, there is a sharp, normatively significant threshold: if at least two of you use the non-toxic paint, the ecosystem will survive; if at most one uses the non-toxic paint, the ecosystem will collapse. This makes *The lake* a typical threshold case. In *Drops of water*, however, it seems that there is no normatively significant threshold: what matters is how the men in the desert feel, and adding one extra pint to the cart makes no perceptible difference to the suffering of anyone, no matter how many others have donated their pints. This feature of *Drops of water* has made it very difficult to find a plausible account of reasons for action that can capture the verdict that you do have a reason to donate your pint.

One way to respond to this challenge is to argue that there is a normatively significant threshold after all. For example, Kagan (2011) argues that non-threshold cases are conceptually impossible (for a reply, see Nefsky 2012), and Parfit (1984), Barnett (2018), and Broome (2019) give arguments aiming to show that imperceptible differences do matter morally. REASON offers an alternative response, which does not depend on how this debate turns out:

According to REASON, what matters is that there is a normatively significant difference between outcomes at opposite ends of the spectrum – for example, between the men's suffering being fully alleviated and their suffering continuing unmitigated. By donating your pint, you can increase the security of the good outcome that the men's suffering is fully alleviated and decrease the security of the bad outcome that

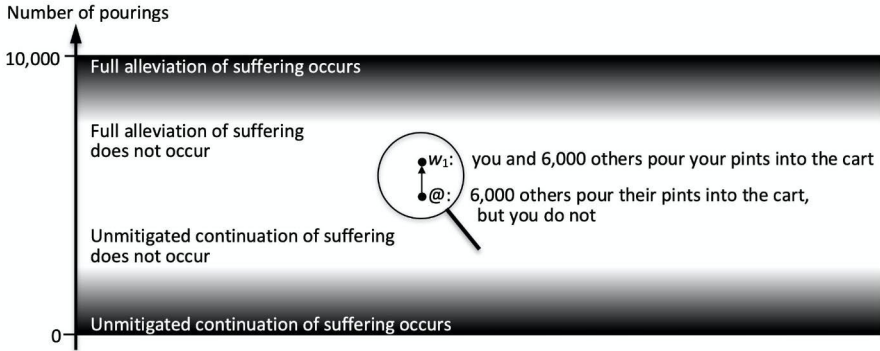


Fig. 5 The figure shows the relevant possibility horizon for determining what you have reason to do in *Drops of water*. Note the vague boundaries between worlds where the men’s suffering is fully alleviated and worlds where it is not, and between worlds where the men’s suffering continues unmitigated and worlds where it does not. In the center, the figure zooms in on two possible worlds: @ where 6,000 others pour their pints into the cart, but you do not; and w_1 where you and 6,000 others pour your pints into the cart

the men’s suffering continues unmitigated. Because of that, you have a reason to donate your pint.²¹

In the following, we show in more detail that the conditions of REASON are satisfied. To do so, we consider a particular situation where 6,000 others donate their pints while you fail to donate yours. Did you have a reason to donate your pint?

By hypothesis, the first two conditions of REASON are satisfied: (a) it is an option for you to donate your pint, and (b) it is an option for you to keep it to yourself. Further, we may choose O and O*, such that

O = full alleviation of the men’s suffering

O* = unmitigated continuation of the men’s suffering

This ensures that condition (c) is satisfied: the full alleviation of the men’s suffering is clearly better than the unmitigated continuation of their suffering. The key here is that O and O* are at opposite ends of the spectrum so that it clearly makes a morally significant difference whether one or the other occurs.

We may now go on to show that (d) is satisfied. To do so, we first need to identify the relevant possibility horizon. You and each of the other 9,999 people can choose between two options: donate your pint or keep it for yourself. The relevant possibility horizon therefore has to contain at least $2^{10,000}$ possible worlds. Of course, we cannot represent all of these worlds individually, but Fig. 5 will hopefully do.

²¹ Kagan, Broome, etc., focus on the two adjacent outcomes that would result if you were to either donate your pint or not. By contrast, our approach is more similar to that of Rabinowicz (1989: 39–43), focusing on the whole spectrum of outcomes – including outcomes that would come about if others acted differently.

To ensure that the distance between @ and w_1 is discernible in Fig.5, we have magnified this part of the figure (as indicated by the stylized magnifying glass). It is vague precisely where the boundary is between worlds where the men's suffering is fully alleviated and worlds where it is not. To illustrate this, we use a gradient that runs from black to white rather than a sharp line. The same applies to the boundary between worlds where the men's suffering continues unmitigated and worlds where it does not.

Intuitively, it is now clear that (d) the full alleviation of the men's suffering is *more secure* in w_1 than it is in @: by adding your pint you *take one step closer* to the full alleviation of the men's suffering. Similarly, the unmitigated continuation of the men's suffering is *less secure* in w_1 than it is in @.

However, one might object that it is not obvious that adding a pint to the cart makes the full alleviation of suffering *more secure* (or the unmitigated continuation of suffering *less secure*): one might argue that since there is no sharp boundary between worlds where the men's suffering is fully alleviated and worlds where it is not, we cannot decide whether the distance between w_1 and the closest-to- w_1 -at- t world(s) with full alleviation is in fact shorter than the distance between @ and the closest-to-@-at- t world(s) with full alleviation – just as we cannot decide the exact distance between a point and an interval with fuzzy boundaries. We have two answers to this worry:

First, even if it is vague precisely where we reach the full alleviation of suffering, it may still be clear that by pouring your pint into the cart, you are taking a step towards the desired outcome. Compare this to being on an airplane that is flying into a cloud: although it is vague precisely when you go from being outside the cloud to being inside the cloud, there is no doubt about whether you are moving *towards* the cloud.

Second and more formally, we may look more closely at vagueness. According to an attractive view – presented, for example, by David Lewis (see also Fine 1975) – vagueness is semantic indecision:

‘The only intelligible account of vagueness locates it in our thought and language. The reason it's vague where the outback begins is not that there's this thing, the outback, with imprecise borders; rather there are many things, with different borders, and nobody has been fool enough to try to enforce a choice of one of them as the official referent of the word “outback”.’ (Lewis, 1986: 213).

We may understand the vagueness of when the men's suffering is fully alleviated in a parallel way: the reason it is vague when the men's suffering is fully alleviated is not that there is this event, the full alleviation of the men's suffering, with imprecise conditions of occurrence; rather there are many events, with different conditions of occurrence, and nobody has been fool enough to try to enforce a choice of one of them as the official referent of ‘the full alleviation of the men's suffering.’

How should we evaluate the truth of a sentence containing vague terms, such as ‘the outback’ or ‘the full alleviation of the men's suffering’? In the following, we will appeal to the supervaluationist proposal (see Fine 1975; Keefe, 2000), which takes the idea that vagueness is semantic indecision as its starting point. To explain the proposal, suppose that there are three areas with completely precise boundaries –

call them Outback_1 , Outback_2 , and Outback_3 . Suppose further that we haven't made a choice whether 'the outback' refers to Outback_1 , Outback_2 , or Outback_3 , but that there are no other competing candidates. The supervenient idea is that, in many cases, it simply does not matter that we haven't decided: if a sentence would be true no matter how we were to decide, then we can say that it is true *without* deciding. Suppose, for example, that the following three fully precise claims are all true: 'David went into Outback_1 ,' 'David went into Outback_2 ,' and 'David went into Outback_3 .' In that case, we can say that the sentence 'David went into the outback' is true without deciding whether 'the outback' refers to Outback_1 , Outback_2 , or Outback_3 : the sentence would come out true no matter how we were to decide.

Introducing a bit of technical language, we may say that a vague term has several admissible completely sharp sharpenings, where an admissible completely sharp sharpening is simply a fully precise term that corresponds to one of the candidate meanings of the vague term (that is, one of the meanings that we have not decided among). For example, ' Outback_1 ' is an admissible completely sharp sharpening of 'the outback.' Then the supervenient proposal goes as follows:

Supervenientism: A statement is supertrue and therefore true if and only if it is true on all its admissible completely sharp sharpenings.

For example, the statement 'David went into the outback' is true on all its admissible completely sharp sharpenings. It is thus supertrue and therefore true.

We may now apply this proposal to *Drops of water*. To do so, we need to identify all admissible completely sharp sharpenings (for short, 'admissible sharpenings') of 'the full alleviation of the men's suffering' and 'the unmitigated continuation of the men's suffering.' We will do this in two steps.

First, we define events with precise conditions of occurrence that consist in the men's experiencing some particular level of suffering. We do this by letting $\text{Exact}(n)$ be the event that occurs just in case each of the ten thousand men experiences the level of suffering he will experience after drinking when *exactly* n pints have been evenly distributed. For example, $\text{Exact}(9,000)$ occurs just in case each of the men experiences the level of suffering he will experience after drinking 0,9 pint.

Second, we use events of the form $\text{Exact}(n)$ as building blocks to characterise admissible sharpenings of 'the full alleviation of the men's suffering' and 'the unmitigated continuation of the men's suffering.' We believe it is plausible that any admissible sharpening of 'the full alleviation of the men's suffering' has the form $\text{AtLeast}(N)$, where $\text{AtLeast}(N)$ is the event that occurs just in case $\text{Exact}(n)$ occurs for some $n \geq N$. For example, $\text{AtLeast}(9,000)$ is one such admissible sharpening of 'the full alleviation of the men's suffering'; it occurs just in case either $\text{Exact}(9,000)$ or $\text{Exact}(9,001)$... or $\text{Exact}(10,000)$ occurs. Similarly, we suggest that any admissible sharpening of 'the unmitigated continuation of the men's suffering' has the form $\text{AtMost}(M)$, where $\text{AtMost}(M)$ occurs just in case $\text{Exact}(n)$ occurs for some $n \leq M$. For example, $\text{AtMost}(1,000)$ is one such admissible sharpening of 'the unmitigated continuation of the men's suffering'; it occurs just in case either $\text{Exact}(1,000)$ or $\text{Exact}(999)$ or ... $\text{Exact}(0)$ occurs.

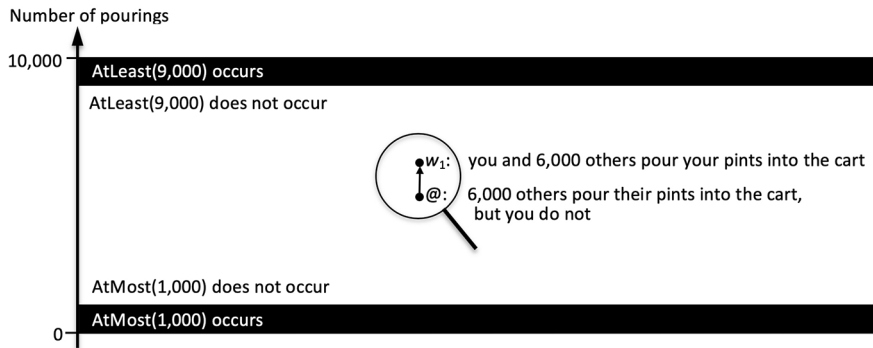


Fig. 6 The figure shows the relevant possibility horizon for determining what you have reason to do in *Drops of water*. Note the sharp boundaries between worlds where *AtLeast(9,000)* occurs and worlds where it does not, and between worlds where *AtMost(1,000)* occurs and worlds where it does not. In the centre, the figure zooms in on two possible worlds: $@$ where 6,000 others pour their pints into the cart, but you do not; and w_1 where you and 6,000 others pour your pints into the cart

We may now show that condition (d) is satisfied for all admissible sharpenings of ‘the full alleviation of the men’s suffering’ and ‘the unmitigated continuation of the men’s suffering.’²² For illustration, we consider the following:

$$O = \text{AtLeast}(9,000)$$

$$O^* = \text{AtMost}(1,000)$$

It is clear that *AtLeast(9,000)* is *more secure* in w_1 than it is in $@$. Starting from w_1 (where you donate your pint), only 2,999 people need to act differently in order for *AtLeast(9,000)* to occur; but starting from $@$, 3,000 people need to act differently. Similarly, *AtMost(1,000)* is *less secure* in w_1 than it is in $@$. Starting from w_1 , 5,001 people need to act differently in order for *AtMost(1,000)* to occur; but starting from $@$, only 5,000 need to act differently. This is illustrated in Fig.6, which indicates the precise conditions of occurrence of *AtLeast(9,000)* and *AtMost(1,000)*.

The same argument applies to any admissible sharpening of ‘the full alleviation of the men’s suffering’ and ‘the unmitigated continuation of the men’s suffering.’ We therefore find that the two statements

‘the full alleviation of the men’s suffering is *more secure* in w_1 than in $@$,’ and
 ‘the unmitigated continuation of the men’s suffering is *less secure* in w_1 than in $@$,’

are supertrue and therefore true, since they are true on any admissible sharpening of ‘the full alleviation of the men’s suffering’ and ‘the unmitigated continuation of the

²² Note that condition (c) is also satisfied for all such admissible sharpenings. For example, it is clearly true that the occurrence of *AtLeast(9,000)* is better than the occurrence of *AtMost(1,000)*.

men's suffering.'²³ REASON thus delivers the desired verdict that you have a reason to donate your pint in a version of *Drops of water* where 6,000 others donate theirs. The same arguments apply no matter how many others donate their pints.

Our treatment of *Drops of water* generalises to show that you have a reason to act in a wide range of collective action problems, including climate change.

8 Understanding why intuitions vacillate

So far, we have seen that REASON delivers the desired verdicts in our test cases and in *Drops of water*. However, a question remains: why do our intuitions vacillate in many of these cases? In this section we show how REASON provides a framework that can explain such vacillation.

You might feel drawn to the verdict that you do not have an objective reason to press your button in *Nuclear safety* or to use the non-toxic paint in *The lake*. If so, you might motivate your verdict by emphasising certain features of the actual situation: as a matter of fact, Suzy *did* press her safety-button; given that she did, why would you have a reason to press yours? Or again, the other two boat owners *did* in fact use the toxic paint; given that they did, the lake's ecosystem was going to collapse no matter what you did – so why would you have a reason to use the non-toxic paint?

REASON and, in particular, the notion of a relevant possibility horizon offers a straightforward way to understand these arguments: when you make these arguments, you are in effect insisting that we should use a smaller possibility horizon – a possibility horizon where it is held fixed that Suzy presses her button or, in the second case, that the other two boat owners use the toxic paint. That is, you are insisting that the relevant possibility horizon simply does not include the possibility that Suzy might not have pressed her button or that each of the other boat owners might have used the non-toxic paint. If we use such a smaller possibility horizon, where we only consider how *you* might have acted differently while holding fixed the actions of everyone else, REASON does indeed deliver the verdict that you had no reason to press your button in *Nuclear safety* and no reason to use the environmentally friendly paint in *The lake*. The restricted possibility horizon in the case of *Nuclear safety* is illustrated in Fig. 7. Within this possibility horizon, there simply is no world with a nuclear disaster. Thus, the safe shutdown of the reactor is infinitely secure no matter whether you press your button or not, and therefore you have no reason to press your button.

If we restrict our possibility horizon in a similar way in the case of *Drops of water*, we similarly find that you have no reason to donate your pint: supposing that 6,000 others in fact donate their pints, the restricted possibility horizon contains only two possible outcomes – Exact(6,000) and Exact(6,001). This is illustrated in Fig. 8.

²³ We would get the same result if we instead appealed to an epistemic view of vagueness according to which vagueness is a kind of ignorance. On such a view, 'the full alleviation of the men's suffering' refers to one event with precise conditions of occurrence; we just don't know which (cf. Williamson, 1994). The same applies to 'the unmitigated continuation of the men's suffering.' Presumably, these events have the form $\text{AtLeast}(N)$ and $\text{AtMost}(M)$. From the reasoning above, it now immediately follows that (c) and (d) are satisfied on an epistemic view.

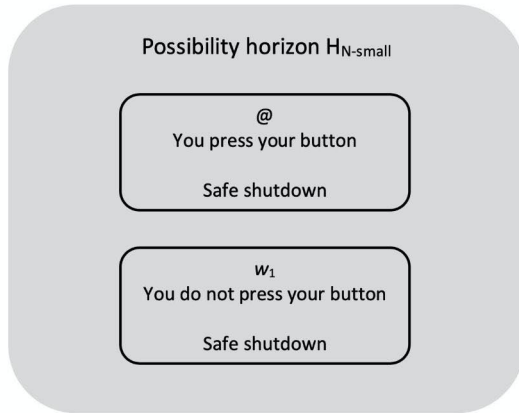


Fig. 7 The figure depicts a restricted possibility horizon in the case of *Nuclear safety*, namely possibility horizon $H_{N\text{-small}}$ with only two worlds: @ where you press your button; and w_1 where you do not press your button. The reactor is shut down safely in both @ and w_1 .

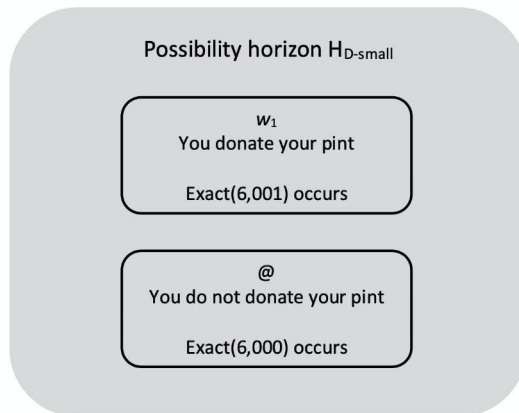


Fig. 8 The figure depicts a restricted possibility horizon in the case of *Drops of water*, namely $H_{D\text{-small}}$ with only two worlds: @ where you do not donate your pint and Exact(6,000) occurs; and w_1 where you donate your pint and Exact(6,001) occurs

By hypothesis, there is no morally significant difference between Exact(6,000) and Exact(6,001): the men experience the same level of suffering. When we treat this restricted possibility horizon as the relevant one, REASON therefore delivers the verdict that you have no reason to donate your pint, since condition (c) is not satisfied.

It seems plausible that intuitions vacillate in some cases because we are torn between treating the smaller or the larger possibility horizon as the relevant one.

This also gives a way to understand disagreements about reasons: when we ostensibly disagree about whether you have a reason (to press your button, use the non-toxic paint, or donate your pint), we may in fact be disagreeing about which possibilities should be included in the relevant possibility horizon. Should we merely include the possibilities corresponding to the actions that are open to *you* (while holding fixed the actions of everyone else), or should we treat it as a relevant possibility that everyone involved in the situation might act differently? Depending on how we answer this question, REASON will deliver different verdicts.

While we recognise the intuitive pull of the smaller possibility horizon, we believe that, in the end, we need to choose the larger possibility horizon in order to get a correct understanding of our reasons: it is only by treating it as a relevant possibility that every agent involved in the situation might have acted differently that we can satisfy *Explaining suboptimal outcomes* and *Explaining optimal outcomes*, and thereby the principle of moral harmony.

9 Conclusion

REASON offers a general account of teleological reasons for action. This account explains *in virtue of what* you have a teleological reason to donate your pint in *Drops of water*: you have such a reason in virtue of the fact that donating your pint makes it *more secure* that the men's suffering will be fully alleviated and *less secure* that the men's suffering will continue unmitigated.

Drops of water brings out the crucial features of many collective action problems, such as climate change. Our account captures the idea that *we together* can make a difference to what happens in such cases. When we together can make a difference, you have a reason to act when your action *contributes* to our making a difference together. Saying that 'you have a reason because we together can make a difference' might evoke top-down accounts of reasons where your individual reasons derive from what the collective can do. However, REASON captures the intuition that 'you have a reason because we together can make a difference' while being entirely bottom-up: our rule for determining the relevant possibility horizon ensures that we consider all the agents involved in a situation and every combination of the courses of action that are open to them. When we apply this to *Drops of water*, we see that the 10,000 can make a significant positive difference: they can ensure that the men's suffering is fully alleviated rather than continuing unmitigated. You contribute to making this difference by making the good outcome *more secure* and the bad outcome *less secure* – that is, by donating your pint rather than keeping it to yourself.

10 Appendix

In this [Appendix](#), we prove four results. We begin with *Stability*: whether you have a reason to ϕ rather than ψ at t does not depend on what you actually do at t . Next, we prove that *The whether-whether inference* and *The causal inference* hold. Finally, we prove *Symmetry*: whenever either O or O* occurs in *every* world within $H(t)$, it is

the case that O is *more secure* in the closest ϕ -ing world than in the closest ψ -ing world if and only if O* is *less secure* in the closest ϕ -ing world than in the closest ψ -ing world.

1. Stability

REASON entails what we call *Stability*:

Stability:

Whether or not you have a reason to ϕ rather than ψ at t does not depend on what you actually do at t .

This is an intuitively pleasing result: it fits the thought that, although you may need to know lots of other things about the actual world in order to figure out whether you have objective reason to ϕ rather than ψ at t , you do not need to know what you will in fact do at t .

Consider REASON. It seems plausible to assume that the range of options you have is independent of which of these options you realize. Thus, the satisfaction of condition (a) and (b) is independent of what you actually do at t . Similarly, if (c) O is better than O*, this is obviously so independently of what you actually do at t . What remains is condition (d), which requires us to look at the closest-to-@-at- t world where you ϕ at t and the closest-to-@-at- t world where you ψ at t . These worlds are the same no matter what you do. To see this, suppose that you have the following range of options: ϕ , ψ , or χ (including further options does not change the structure of the argument). Since ϕ , ψ , or χ are all of your options, you do one of these in the actual world. It now follows from our construction of the relevant possibility horizon $H(t)$ that $H(t)$ includes three worlds – w_ϕ , w_ψ , and w_χ – that are exactly alike, except that you do ϕ in w_ϕ , ψ in w_ψ , and χ in w_χ . One of these three worlds is @. Irrespective of what you do – that is, irrespective of which of the three worlds is @ – we find that the closest-to-@-at- t world where you ϕ is w_ϕ , and the closest-to-@-at- t world where you ψ is w_ψ . It therefore makes no difference to the verdict of REASON whether you do ϕ , ψ , or χ .

2. The whether-whether inference

REASON entails that, when you are in a situation where condition (a), (b), and (c) of REASON are satisfied, the following holds:

The whether-whether inference:

If *whether* O or O* will occur depends on *whether* you ϕ or ψ at t , then you have a reason to ϕ rather than ψ at t .

More carefully, the antecedent is satisfied when the following holds: in the closest world(s) in $H(t)$ where you ϕ at t , O occurs; and in the closest world(s) in $H(t)$ where you ψ at t , O* occurs.

Suppose that this is the case. Since O and O* are incompatible, O does not occur in the closest ψ -ing world(s), and O* does not occur in the closest ϕ -ing world(s). Thus, O has positive security in the closest ϕ -ing world(s) and negative security in

the closest ψ -ing world(s). Correspondingly, O^* has negative security in the closest ϕ -ing world(s) and positive security in the closest ψ -ing world(s). It immediately follows that (d) O is *more secure* and O^* is *less secure* in the closest ϕ -ing world(s) than they are in the closest ψ -ing world(s). Since condition (a), (b), and (c) are also satisfied, REASON yields the verdict that you have a reason to ϕ rather than ψ at t .

3. The causal inference

REASON also entails that, when you are in a situation where condition (a), (b), and (c) of REASON are satisfied, the following holds:

The causal inference:

If your ϕ -ing rather than ψ -ing at time t would be a cause of O rather than O^* , then you have a reason to ϕ rather than ψ at t .

More carefully, the antecedent is satisfied when the following holds: in the closest world(s) in $H(t)$ where you ϕ rather than ψ at t , your ϕ -ing rather than ψ -ing at time t is a cause of O rather than O^* within $H(t)$.

To prove that *The causal inference* holds, we need a result about causation, namely that condition (d) of REASON expresses a necessary condition for causation (see Sect. 5):

Your ϕ -ing rather than ψ -ing at time t is a cause of O rather than O^* within $H(t)$ *only if* O is *more secure* and O^* is *less secure* at t in $H(t)$ in the closest-to-@-at- t world(s) in $H(t)$ where you ϕ at t than they are in the closest-to-@-at- t world(s) in $H(t)$ where you ψ at t .

The argument then goes as follows:

Suppose first that you ϕ at t in @. Then @ is the closest world where you ϕ rather than ψ at t . If the antecedent is satisfied, it follows that your ϕ -ing rather than ψ -ing is a cause of O rather than O^* within $H(t)$. Given that condition (d) is a necessary condition for causation, it follows that condition (d) is satisfied. Thus, all four conditions of REASON are satisfied, yielding the verdict that you have a reason to ϕ rather than ψ at t .

Suppose next that you do not ϕ at t in @. Let w_ϕ be the closest world in $H(t)$ where you ϕ at t . If the antecedent is satisfied, it follows that in w_ϕ your ϕ -ing rather than ψ -ing is a cause of O rather than O^* within $H(t)$. Given that condition (d) is a necessary condition for causation, it follows that condition (d) is satisfied in w_ϕ . By the same argument as in *Stability*, it now follows that condition (d) is also satisfied in @. Thus, all four conditions of REASON are satisfied, yielding the verdict that you have a reason to ϕ rather than ψ at t .

4. Symmetry

Finally, we show that REASON entails what we call *Symmetry*:

Symmetry:

Suppose that either O or O* occurs in *every* world within $H(t)$. Then the following holds for any two worlds w_1 and w_2 within $H(t)$: O is *more secure* in w_1 than it is in w_2 if and only if O* is *less secure* in w_1 than it is in w_2 .

Consider an arbitrary world w within $H(t)$. Then either O or O* occurs in w while the other does not (since they are incompatible). Suppose that O occurs in w . Then O has positive security in w , and its degree of positive security is given by the distance-at- t between w and the closest-to- w -at- t world(s) in $H(t)$ where O does not occur. O* has negative security in w , and its degree of negative security is given by the distance-at- t between w and the closest-to- w -at- t world(s) in $H(t)$ where O* occurs. Since either O or O* occurs in *every* world within $H(t)$, the closest-to- w -at- t world(s) in $H(t)$ where O* occurs *just are* the closest-to- w -at- t world(s) where O does not occur. The distance that determines O's degree of positive security in w therefore also determines O*'s degree of negative security in w . A parallel argument applies if we suppose that O* occurs in w . From this, the result follows.

Acknowledgements We would like to thank Gunnar Björnsson and Samuel Lee for their close reading and valuable comments. We are also grateful to audiences at the Higher Seminar in practical philosophy at Lund University, at the workshop on Collective and Shared Responsibility at the MANCEPT workshops in political theory at Manchester University, and at the Higher Seminar in philosophy at Umeå University. We would especially like to thank Per Algander, David Alm, Henrik Andersson, Olle Blomberg, Stephanie Collins, Dan Egonsson, Anton Emilsson, Nils Franzén, Kalle Grill, Niels de Haan, Sofia Jeppsson, Ingvar Johansson, Marta Johansson Werkmäster, Christian Löw, Luise Mirow, Ethan Nowak, Björn Petersson, Marcel Quarfood, Wlodek Rabinowicz, Toni Rønnow-Rasmussen, Lars Samuelsson, David Shoemaker, Pär Sundström, Matt Talbert, Bram Vaassen, Jakob Werkmäster, Bill Wringe, and Martín Abreu Zavaleta. Finally, we would like to thank two anonymous reviewers for helping us to develop and clarify many aspects of the paper.

Funding Open access funding provided by Umea University.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barnett, Z. (2018). No free lunch: The significance of tiny contributions. *Analysis*, 78(1), 3–13. <https://doi.org/10.1093/analys/anx112>
- Björnsson, G. (2011). Joint responsibility without individual control: Applying the explanation hypothesis. In J. van den Hoven, I. van de Poel, & N. Vincent (Eds.), *Moral responsibility: Beyond free will and determinism* (pp. 181–199). Springer.
- Björnsson, G. (2014). Essentially shared obligations. *Midwest Studies In Philosophy*, 38(1), 103–120. <https://doi.org/10.1111/misp.12019>

- Braham, M., & van Hees, M. (2012). An anatomy of moral responsibility. *Mind*, 121(483), 601–634. <https://doi.org/10.1093/mind/fzs081>
- Broome, J. (2019). Against denialism. *The Monist*, 102(1), 110–129. <https://doi.org/10.1093/monist/ony024>
- Fanciullo, J. (2020). What is the point of helping? *Philosophical Studies*, 177, 1487–1500. <https://doi.org/10.1007/s11098-019-01263-7>
- Feldman, F. (1980). The principle of moral harmony. *Journal of Philosophy*, 77(3), 166–179. <https://doi.org/10.2307/2025668>
- Fine, K. (1975). Vagueness, truth and logic. *Synthese*, 30, 265–300. <https://doi.org/10.1007/BF00485047>
- Hindriks, F. (2022). The problem of insignificant hands. *Philosophical Studies*, 179, 829–854. <https://doi.org/10.1007/s11098-021-01696-z>
- Kagan, S. (2011). Do I make a difference? *Philosophy & Public Affairs*, 39(2), 105–141. <https://doi.org/10.1111/j.1088-4963.2011.01203.x>
- Keefe, R. (2000). *Theories of vagueness*. Cambridge University Press.
- Lewis, D. (1986). *On the plurality of worlds*. Blackwell.
- Nefsky, J. (2012). Consequentialism and the problem of collective harm: A reply to Kagan. *Philosophy & Public Affairs*, 39(4), 364–395. <https://doi.org/10.1111/j.1088-4963.2012.01209.x>
- Nefsky, J. (2017). How you can help, without making a difference. *Philosophical Studies*, 174, 2743–2767. <https://doi.org/10.1007/s11098-016-0808-y>
- Norcross, A. (2005). Harming in context. *Philosophical Studies*, 123, 149–173. <https://doi.org/10.1007/s11098-004-5220-3>
- Parfit, D. (1984). *Reasons and persons*. Clarendon Press.
- Paul, L. A., & Hall, N. (2013). *Causation: A user's guide*. Oxford University Press.
- Pettit, P. (2015). *The robust demands of the good*. Oxford University Press.
- Pinkert, F. (2015). What if I cannot make a difference (and know it). *Ethics*, 125(4), 971–998. <https://doi.org/10.1086/680909>
- Portmore, D. (2018a). Teleological reasons. In D. Star (Ed.), *The Oxford handbook of reasons and normativity* (pp. 765–782). Oxford University Press.
- Portmore, D. (2018b). Maximalism and moral harmony. *Philosophy and Phenomenological Research*, 96(2), 318–341. <https://doi.org/10.1111/phpr.12304>
- Rabinowicz, W. (1989). Act-utilitarian prisoner's dilemmas. *Theoria*, 55(1), 1–44. <https://doi.org/10.1111/j.1755-2567.1989.tb00720.x>
- Regan, D. (1980). *Utilitarianism and co-operation*. Clarendon Press.
- Schaffer, J. (2012). Causal contextualism. In M. Blaauw (Ed.), *Contrastivism in philosophy* (pp. 35–63). Routledge.
- Schroeder, M. (2007). *Slaves of the passions*. Oxford University Press.
- Singer, P. (1980). Utilitarianism and vegetarianism. *Philosophy & Public Affairs*, 9(4), 325–337
- Sinnott-Armstrong, W. (2008). A contrastivist manifesto. *Social Epistemology*, 22(3), 257–270. <https://doi.org/10.1080/02691720802546120>
- Skorupski, J. (2010). *The domain of reasons*. Oxford University Press.
- Snedegar, J. (2017). *Contrastive reasons*. Oxford University Press.
- Streumer, B. (2007). Reasons and impossibility. *Philosophical Studies*, 136, 351–384. <https://doi.org/10.1007/s11098-005-4282-1>
- Touborg, C. (2018). *The dual nature of causation: Two necessary and jointly sufficient conditions*. PhD thesis, University of St Andrews. <https://research-repository.st-andrews.ac.uk/handle/10023/16561>
- Williamson, T. (1994). *Vagueness*. Routledge.