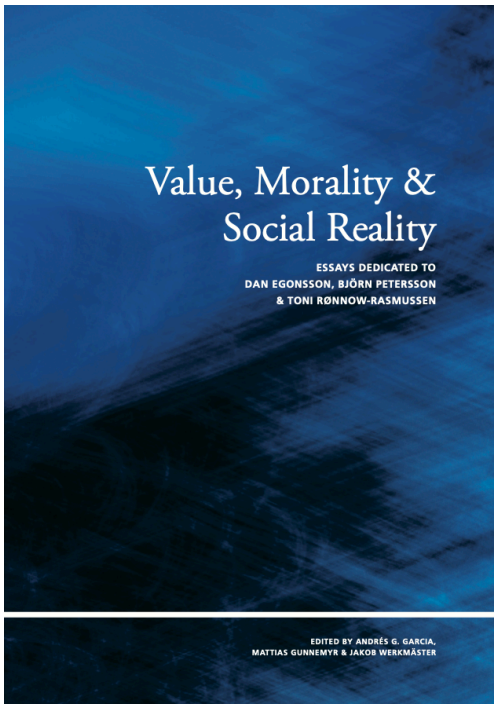
Harming Others

Mattias Gunnemyr

In: Garcia, A., Gunnemyr, M. & Werkmäster, J. (2023) *Value, Morality & Social Reality: Essays dedicated to Dan Egonsson, Björn Petersson & Toni Rønnow-Rasmussen.* Lund: Department of Philosophy, Lund University. DOI: https://doi.org/10.37852/oblu.189

ISBN: 978-91-89415-65-2 (print), 978-91-89415-66-9 (digital)



Published by the Department of Philosophy, Lund University.
Edited by: Andrés Garcia, Mattias Gunnemyr, and Jakob Werkmäster
Cover image by Fabian Jones. Cover layout by Gunilla Albertén.

DOI: https://doi.org/10.37852/oblu.189.c521

This text is licensed under a Creative Commons Attribution-NonCommercial license.
This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license does not allow for commercial use.
(License: http://creativecommons.org/licenses/by-nc/4.0/)

Text © Andrés Garcia, Mattias Gunnemyr, and Jakob Werkmäster 2023.
Copyright of individual chapters is maintained by the chapters' authors.

Harming Others

Mattias Gunnemyr

Abstract. There are two standard accounts of what it is to harm someone: the counterfactual comparative account and the non-comparative account. The first gives counterintuitive verdicts in cases of overdetermination and pre-emption, and the second implausibly entails that an event might harm you even though it makes you better off. On some interpretations, the non-comparative account also gives counterintuitive verdicts in switching cases. I suggest that we can combine these accounts in a way that avoids giving the mentioned counterintuitive implications. Roughly, the suggestion is that you harm someone, *S,* if (a) there is a genuine process connecting what you did to *S*'s being worse off (a non-comparative condition), and (b) *S*'s being better off would have been more secure if you had not acted in this way (a counterfactual comparative condition).

Last week, this paper was rejected by the *Journal of Over-Determination.* According to the journal's strict policy, manuscripts are rejected if at least one of the reviewers recommend rejection and accepted if both reviewers recommend accepting the paper. Deeming from the comments I got, both reviewers considered the paper to be of high quality. However, in the end, both decided to recommend rejection. The review process was double-blinded, but out of a coincidence I later learned that it was reviewed by Björn and Dan.

I would have been better off if the paper had been accepted. I would have been happy that my paper finally found its home in a highly ranked philosophy journal, and my chances of getting tenure would have been improved. Still, it does not matter for me whether one or two of the reviewers recommended rejection. The consolation

of having one positive review would have been balanced out by the frustration of almost getting the paper published.[1]

In recommending rejection, did Björn and Dan harm me? We seem to lack a clear intuitive verdict about whether they did. On the one hand, it seems they did. After all, if neither of them had suggested to reject my paper, it would have been accepted and I would have been better off. On the other hand, it seems that neither of them harmed me. I would not have been better off had Björn recommended to accept my paper since Dan would have recommended to reject it anyway. And similarly, I would not have been better off had Dan recommended to accept my paper since Björn would have recommended to reject it anyway.

The Counterfactual Comparative Account of Harming

Can our best theories of what it is to harm someone explain these shifting intuitions? At a first glance, it seems that they cannot. Consider what I take to be the standard view in the literature on harming:

> *The counterfactual comparative account of harming (CCA):* An event *C* harms a person *S* if and only if *S* would have been better off if *C* had not occurred.
>
> (Feinberg 1984; Parfit 1984; Norcross 2005; Bradley 2009; Klocksiem 2012)

CCA is usually taken to be an analysis of when an event *C* is *all-things-considered* harmful for *S;* that is, when an event *C* makes *S* have a lower lifetime well-being level. On this reading, applying CCA to some situation might require quite laborious evaluations. For instance, in *Journal of Over-Determination,* it requires us to consider whether I as a result of Björn's and Dan's recommendations eventually would get the paper published in another even more well-regarded journal and the wellbeing level I would have then, or whether I would have to end my career in philosophy and pursue some other career entirely, and the wellbeing level that would provide me. To keep the discussion focused, I will instead take CCA to be an analysis of what it is for an event *C* to be *pro tanto* harmful for *S*; that is, what it is for *C* to be harmful for *S* in some respect. This allows me for instance to consider the question of whether Björn's and Dan's recommendations harmed me in the sense of making me sad and having fewer chances of getting tenure without considering whether their recommendations also harmed me or benefitted me in some other way. Still, the arguments I give here could be amended to apply to the standard reading of CCA using slightly modified examples, so using this non-standard interpretation of CCA should not make any difference for the conclusions I draw.

[1] I will call this case *Journal of Over-Determination.* The case is a modified version of an example discussed by Petersson (2018). For further discussion of this example, see e.g. Johansson (this volume).

According to CCA, then, neither Björn's nor Dan's recommendation to reject my paper harmed me. Because of Dan's recommendation, I would not have been better off (happy and with better chances of getting tenure) than I am now (sad and with lesser chances of getting tenure) had Björn not decided to recommend rejection, and because of Björn's recommendation, I would not have been better off had Dan not recommended rejection. Thus, CCA can straight-forwardly explain one of the intuitions we might have about the case: the intuition that Björn's recommendation to reject did not harm me since the paper would have been rejected anyway (and that a similar thing could be said about Dan's recommendation). Yet, this verdict seems less than satisfactory. We have not yet found an explanation for the other intuition: that I was harmed as a result of their recommendations.

Upon closer scrutiny, however, there is a close cousin to CCA that seems to go some way to explain this intuition. This alternative principle does not apply to events, but to *sets of events* (Parfit 1984) or *pluralities* (Feit 2015, 2016). A plurality can be understood as several individual events taken together, so when a plurality harms someone, there are several events such that they harm this person.[2]

> *The plural harm principle (PH):* A set of events – or a plurality – *C* harms a person *S* if and only if *S* would have been better off if *C* had not occurred.[3]

According to Feit and Parfit, PH entails that the two recommendations harmed me: Had not this set or plurality occurred, I would have been better off. That is, had not Björn's *and* Dan's recommendations to reject occurred, they would both have recommended the journal to accept my paper, and I would have been better off. Moreover, PH retains some of CCA's implications: it entails that neither Björn's nor Dan's recommendation harms me. It is still the case that had Björn's recommendation to reject (we can think of this as a one-event plurality) not occurred I would not have been better off, and the same goes for Dan's recommendation.

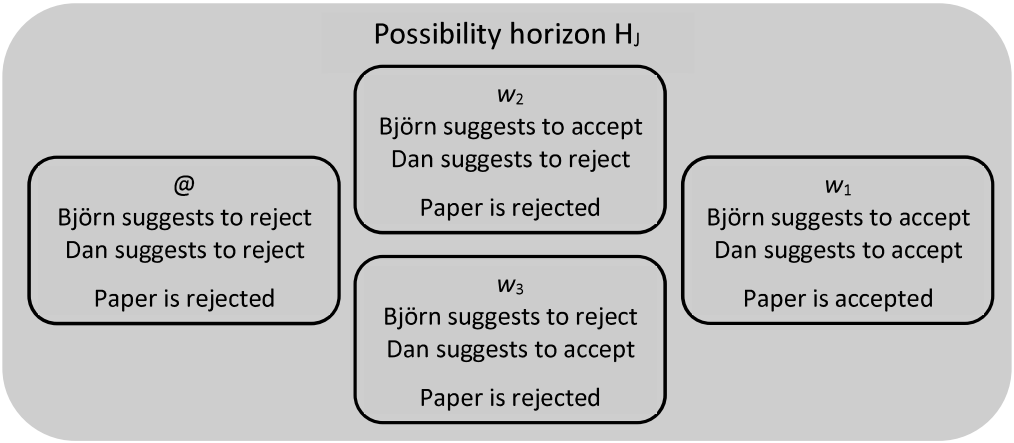
Still, it is far from clear that Parfit and Feit are correct that PH implies that the set or plurality consisting of Björn's and Dan's recommendations harmed me. As Petersson (2004, 2018) argues, when we are deciding whether someone would have been better off had a certain set of events not occurred, we should consider what happens in the closest possible world where this set does not occur. In a case like *Journal of Over-Determination* in which the relevant actions are counterfactually independent of each other, the closest possible world where the set consisting of Björn's and Dan's recommendations does not occur is a world where one of these recommendations still occurs. It might be the world where Björn but not Dan

[2] Feit (2015) does not make entirely clear what a plurality is. For discussion, see Petersson (2018: 846-7).

[3] Parfit (1984) and Feit (2015) place a further restriction on what it is for a set (or plurality) of events to harm someone: A set of events harms A if and only if that set is *the smallest set* such that, if none of the events had occurred, A would have been better off. They include the extra restriction to avoid the contra-intuitive result that completely unrelated events, such as Fred Astaire's dancing in the distance, also belong to a set that harmed me. Here, I will set aside this complication.

recommends rejection, or the world where Dan but not Björn recommends rejection. Hence, if the set consisting of Björn's and Dan's recommendations had not occurred, one of these recommendations would still have occurred, and I would not have been better off than I would have been if both had recommended rejection. So, while it might seem that PH entails that Björn's and Dan's recommendations together harmed me, this verdict is upon closer reflection mistaken.[4]

This point might need some elucidation. To get any counterfactual analysis to deliver plausible results, we need to decide which possible worlds are relevant. As Norcross (2005), Klocksiem (2012) and others point out, the relevant worlds are often provided by context. In the case under consideration, it is not made explicit what would happen in the closest possible world where Björn does not recommend rejection. It is assumed from context, I take it, that he would recommend accepting the paper. Still, upon reflection, he might do other things. He might for example contact Dan and persuade him that the paper deserves to be published, with the result that they both recommend the journal to accept the paper. In that case, CCA and PH would not have the implications we thought they had. They would entail that Björn's recommendation to reject harmed me: I am better off in the closest possible world where he does not recommend rejection (the world where he contacts Dan). This open-endedness will not do. To be able to evaluate any counterfactual analysis, we need to decide which possible worlds are relevant. Petersson's argument (and Parfit's and Feit's arguments, for that matter) presupposes that we have done so already. In *Journal of Over-Determination,* for instance, it is implicitly assumed that there are four relevant possible worlds: the actual world (@) where both Björn and Dan recommends rejecting the paper, the possible world where none of them recommends rejection (w_1), and the possible worlds where one of them but not the other suggest rejection (w_2 and w_3). This is the *possibility horizon* relevant for the case as Touborg (2018) would say, inspired by what Mackie (1974) calls a "causal field".



[4] Like Parfit and Feit, Jackson (1997) also mistakenly concludes that the relevant set of actions is harmful in cases of overdetermined harm.

Note that this possibility horizon does not include possibilities not alluded to in the original description of the case. It does for instance not include the possible world where Björn contacts Dan and persuades him to recommend the journal to accept my paper.

Petersson's argument implicitly assumes a possibility horizon like H_J, and it can be reconstructed as follows: Since Björn's and Dan's recommendations are counterfactually independent of each other, the relevant possible world closest to @ is not w_1, but w_2 or w_3, depending on whether Björn or Dan was closer to recommending the journal to accept the paper. Therefore, my paper had been rejected even if not both had suggested to reject it. It had been rejected since one of them still had recommended rejection. We can then conclude that PH applied to this possibility horizon yields the counter-intuitive result that the set consisting of Björn's and Dan's recommendations did not harm me. We can also conclude that while PH can explain the intuition that Björn's recommendation did not harm me since my paper would have been rejected anyway (and that an analogous thing may be said about Dan's recommendation to reject), it fails to explain the intuition that their recommendations harmed me – just like CCA does.

A defender of the idea that the set or plurality consisting of Björn's and Dan's recommendations harms me might argue that the relevant comparison is not the one between the actual world where both of them recommend rejection (@) and the world in which only one of them does (w_2 or w_3), but the one between the actual world and the world in which none of them does (w_1). This is the position that Parfit and Feit take. Parfit states that we should consider what happens "if they had all acted differently" (1984: 71) and Feit likewise says that we should consider what happens "if none of [the relevant events] had occurred" (2015: 371).

This does not help much. First, as Petersson (2018) points out, "the relevant counterfactual comparison is given by the case-description" (846), and if the principle we use requires that we make another counterfactual comparison than the one given by the case, it is not applicable. In *Journal of Over-Determination,* it is given by the case-description that in the closest relevant possible world where Björn does not recommended rejection, Dan still does (and vice versa), which entails that in the closest relevant possible world where the set consisting of two recommend-dations to reject does not occur, one of the reviewers still recommends rejection. The requirement that we take the world where they both act differently as the relevant one makes PH inapplicable to this case, or at least – we might add –insists that we ignore essential features of the case.

Second, even if we agree with Parfit and Feit that the relevant possible world for comparison is the one where neither Björn nor Dan recommend rejection, and that this principle is applicable to the cases like *Journal of Over-Determination,* we still end up in an unattractive position. While we get the result that the set consisting of their recommendations harmed me, we also get the result that neither Björn's nor Dan's recommendation harmed me. Neither Parfit nor Feit take this to be an unattractive position. Parfit says that it might still be wrong of me to perform some

act if this act is part of a harmful set of acts even though my act harms no one. Feit says (considering structurally similar cases) that while we cannot say that Björn's recommendation or Dan's recommendation harmed me, we can still say that each of them was *involved* in harming me. Still, it seems attractive to be able to say that Björn's recommendation was one the acts that harmed me, and likewise for Dan's recommendation. After all, each recommendation was sufficient for bringing this outcome about.[5]

In a way, Petersson (2018) takes one step further than Parfit and Feit do. He argues that we should hold on to CCA and conclude that neither Björn's recommendation, Dan's recommendation, nor the set of events consisting of both recommendations harmed me.[6] This conclusion seems counterintuitive, and I think we can do better than that.

A Better Account

If we consider *Journal of Over-Determination* more closely, we see that my paper is closer to getting accepted in the possible worlds w_2 and w_3 where either Björn or Dan suggests accepting my paper than in the actual world @ where both of them suggest rejecting it. I am not better off in these other worlds than in the actual one – the consolation of having one positive review is balanced out by the frustration of almost getting the paper published – but there is a difference in terms of closeness to the world where I am better off. We can say that the rejection of my paper is more *secure* when both Björn and Dan recommend rejection, than it is when only one of them does. Similarly, the acceptance of my paper is more secure – it is closer to happening – in the worlds where only one of them recommends rejection than it is in the world where both do. We can use this insight to construct a more accurate analysis of what it is to harm someone:

> *The security comparative account of harming (SCA):* An event *C* harms a person in H if and only if *S*'s being better off in this respect would have been more secure in H had *C* not occurred.[7]

Here, H stands for "possibility horizon". In relativising what it is for an event to harm someone to H, this principle makes explicit the idea that we must specify the relevant possible worlds before we apply our analysis.

---

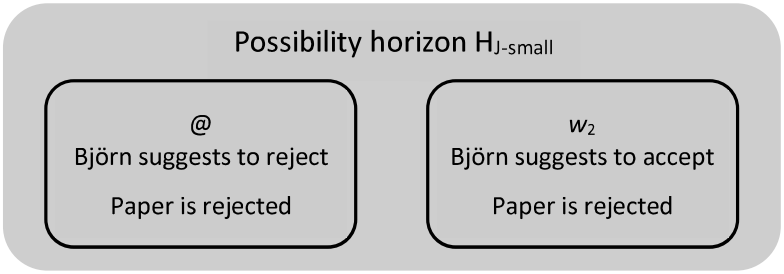
[5] PH faces other problems as well. See e.g. Johansson and Risberg (2019) and Johansson (this volume), especially his discussion on premise 1.

[6] Petersson (2018: 848-9) gives some reasons why this seemingly counterintuitive position might not be so counterintuitive after all. For brevity, I omit discussing these reasons.

[7] In this paper, SCA is meant to evaluate whether an event *C* is *pro tanto* harmful for *S*. However, the arguments I give here could be amended to apply to a reading of SCA in terms of overall harm.

SCA entails that Björn's recommendation to reject the paper harmed me, given H_J. Even though I would not have been better off in the closest-to-@ world where Björn did not recommend rejecting it, my being better off is more secure had he not recommended rejection. Likewise, given H_J, SCA entails that Dan's recommendation to reject harmed me. Moreover, if we allow for SCA to be applied to sets (or pluralities) of events, it also entails that the set consisting of Björn's and Dan's rejection harmed me. Had this set not occurred, I would not have been better off, but my being better off had been closer to occurring. Even without allowing for SCA to be applied to sets (or pluralities), we might still say that there is a set (or plurality) of events that harms me: each of Björn's and Dan's recommendations harms me, and together they constitute a set of events where the each event in that set individually harms me.

SCA can also explain our shifting intuitions about whether Björn and Dan harmed me. As argued above, if we assume that H_J captures the relevant possibilities in the case, SCA entails that each of the recommendations harmed me, and possibly that the set of recommendations harmed me. However, there is another way of understanding the case. We might think that I would not have been better off had Björn decided to suggest the journal to accept my paper since Dan decided to reject it anyway. That is, given that Dan decided to reject the paper, Björn's decision does not matter for whether the paper is rejected or not. There is a natural reading of this idea in terms of which possibilities that are relevant. When we say things like "given that Dan decided to reject the paper", or "Dan decided to reject it", it seems that we are not treating it as an open possibility that Dan could have acted otherwise. That is, we seem to treat Dan's decision as a background condition rather than a potential cause of harm. If we do, we do not really treat the case as involving four relevant possibilities, but only two: that Björn either decides to suggest rejection or decides to suggest the journal to accept the paper. As a result, we get a much smaller possibility horizon, as follows:



When applied to this smaller possibility horizon, SCA entails that Björn's recommendation did not harm me. The rejection of the paper is as secure in w_2 as it is in @. This point might need some elaboration. We can think about the degree of

security in terms of distances between possible worlds. The security of an outcome in a certain possible world is the distance between this world and the closest possible world where this outcome does not occur. To decide the security of the outcome that the paper is rejected in @, we thus have to look at the distance between this world and the closest-to-@ world in $H_{J\text{-small}}$ where this outcome does not occur. However, there is no such world in $H_{J\text{-small}}$. If we start out from @ and travel out into the modal space containing only relevant possible worlds, we will never encounter a world where the paper is not rejected. Therefore, we might say that the distance in question is infinite, and that the outcome that the paper is rejected in @ is infinitely secure. The same goes for the security of the outcome that the paper is rejected in w_2. Hence, the security of the outcome that the paper is rejected is the same in @ as it is in w_2, namely infinitely secure.

We can thus understand the two intuitions as stemming from different possibility horizons. The intuition that Björn and Dan harmed me originates in the larger possibility horizon where we treat it as an open possibility that one or both of them could have recommended otherwise, and the intuition that neither of them harmed me originates in smaller possibility horizons such as $H_{J\text{-small}}$. If we take it as a given that Dan will recommend rejecting my paper, it seems that Björn's recommendation did not matter for whether I was harmed or not. Likewise, if we treat Björn's recommendation as a background condition, Dan's recommendation seems to make no difference for whether I am harmed or not.[8]

This helps us see where CCA goes wrong (which it does, *pace* Petersson 2018). On a standard understanding, CCA tells us to compare the actual world with the closest-to-@ world where *C* does not occur. In effect, it tells us to only take into consideration possibility horizons such as $H_{J\text{-small}}$, making us blind to the fact that there are worlds further away that might be relevant in the evaluation. Granted, CCA (on its standard interpretation) might allow for considering larger possibility horizons such as H_J. It does not exclude the idea that there is a possibility that for example both Björn and Dan would recommend accepting my paper. Still, even if we take such larger possibility horizons into consideration, CCA promptly tells us that the comparison relevant for determining whether *C* harms *S* is the one between the actual world @ and the closest-to-@ world where *C* does not occur. Thereby, possible worlds further away will never matter in the evaluation. Yet, if we allow them to matter contrary to the recommendations of CCA, we find plausible ways of explaining our intuitions in cases like *Journal of Over-Determination.* Similar remarks apply to PH.

[8] Norcross (2005) makes a similar suggestion. He argues that the conversational context indicates the relevant possible world for evaluating whether an event harms someone, and so that an event might be correctly described as harming in one conversational context, and correctly described as benefitting in another.

Pre-emption Cases

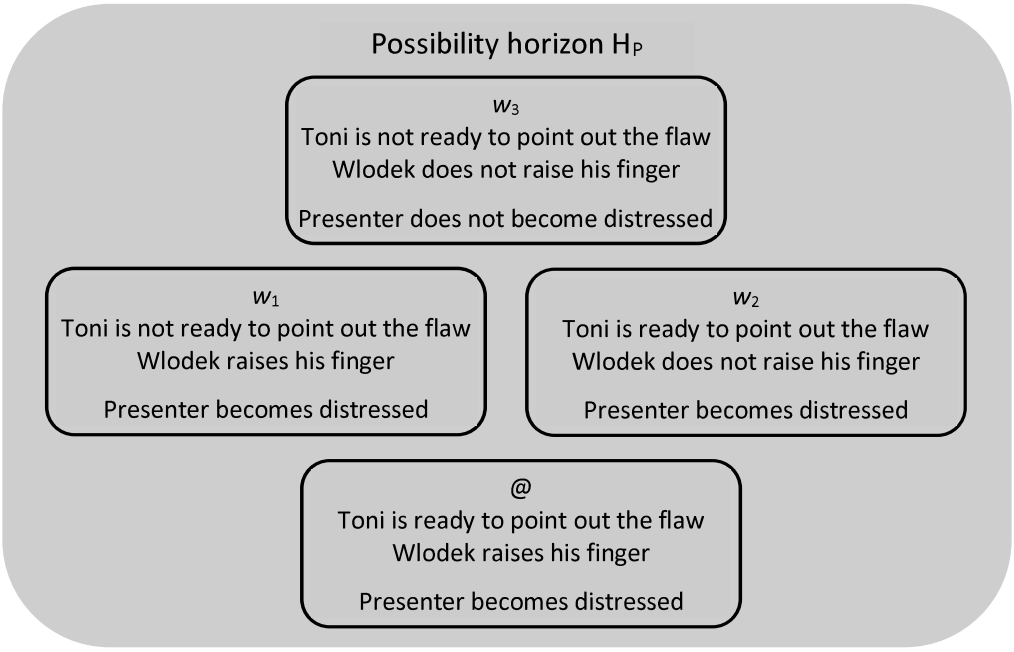
While SCA gives intuitively correct verdicts about harming in some notoriously tricky cases, it cannot be the accurate account of harming we are looking for. It gives patently erroneous verdicts in pre-emption cases, such as the following:

> *Presentation on pre-empted harm:* Toni and Wlodek are listening to a presentation on the topic pre-empted harm at the Higher Seminar. At the same time, they spot a crucial flaw in the argument. Immediately, Wlodek raises his index finger to signal to the chair that he wants to say something. Wlodek gets the word and explains the crucial flaw in the argument with the result that the presenter becomes quite distressed. Had Wlodek not raised his finger, Toni would have raised his a few moments later and explained the flaw in the argument with the result that the presenter had been just as distressed. However, when seeing Wlodek raising his finger, Toni decided not to raise his. He thought that one interruption to the presentation was enough.

Here, it seems clear that Wlodek but not Toni harmed the presenter. You might object that pointing out crucial flaws in someone's argument is not harmful but rather beneficial for this person. Doing so helps her to improve her arguments, to discard mistaken theses, and to make philosophical progress. This objection is relevant, but rests on a misunderstanding. SCA is meant to capture what it is for an event to be *pro tanto* harmful for someone, or in other words, to be harmful *in some respect.* Bearing this in mind, it seems plausible to say that Wlodek (but not Toni) harmed the presenter in one respect: Wlodek's comment made the presenter distressed during the presentation. We could agree that Wlodek's comment also benefitted the presenter in several ways, we could even agree that Wlodek's comment all-things-considered was beneficial for the presenter, and still hold on to the idea that Wlodek's comment harmed the presenter in this one way. In fact, I think this is precisely how we should understand the situation.

Now, it is clear that Wlodek but not Toni harmed the presenter in a way. Toni would have harmed the presenter had Wlodek not beaten him to it, but as things turns out, he did not. SCA, however, entails that both Wlodek and Toni harmed the presenter. To see this properly, we first have to settle the relevant possibility horizon. Consider the time at which Wlodek raises his index finger. At this time, there are four relevant possibilities, as indicated in the possibility horizon H_P (depicted on the next page).

Applied to H_P, SCA wrongly entails that Toni harmed the presenter. In the closest-to-@ world where Toni is not ready to point out the crucial flaw in the argument (w_1), the presenter still becomes distressed. However, this outcome is less secure than in the actual world. In w_1, the only thing that needs to change for the presenter not to become distressed is Wlodek raising his finger, wanting to say



something, whereas in @, Wlodek raising his finger *and* Toni's readiness to point out the flaw need to change in order for the presenter not to become distressed.
For comparison, CCA also gives inaccurate verdicts in pre-emption cases like this. While SCA entails that both Toni and Wlodek harmed the presenter, CCA entails that neither Toni nor Wlodek harmed the presenter. Toni does not harm the presenter since, even if he had not been ready to point out the flaw, Wlodek would have pointed out the flaw anyway. Wlodek, in turn, did not harm the presenter according to CCA since, had he not raised his finger and pointed out the flaw, Toni would have done so instead, and the presenter had become just as distressed anyway.[9]

Non-comparative Accounts of Harm

Seeing that comparative accounts of harming like CCA, PH and SCA run into trouble, we might be tempted to turn to non-comparative accounts, such as the following:

[9] This problem has been pointed out and discussed by e.g. Bradley (2012) and Johansson and Risberg (2019).

> *Non-comparative account of harming (NCA):* An event harms someone if it causes the person to be in an intrinsically bad state.
>
> (see e.g. Shiffrin 1999; Harman 2009)

If we take what it is for an event to cause another at face value, NCA seems to give the right verdicts in the cases we have considered so far. It is Wlodek and not Toni who causes the presenter to be distressed, and to be distressed is an intrinsically bad state. So, NCA entails that Wlodek but not Toni harms the presenter (in this respect). Further, both Björn's and Dan's recommendations caused my article to be rejected, and by extension they caused my sadness and my impoverished chances of getting tenure (two intrinsically bad states, we might assume). Therefore, NCA entails that both Björn's and Dan's recommendations harmed me.

We might consider what kind of account of causation that would fit NCA. For a start, a simple counterfactual account of causation – sometimes called the But-For test for causation – reinvites trouble. According to this account, *C* causes *E* if and only if had *C* not occurred, *E* had not occurred.[10] It entails that neither Björn's nor Dan's recommendation caused my article to be rejected in *Journal of Over-Determination*, and that neither Wlodek nor Toni caused the presenter distress in *Presentation on pre-empted harm.* As a result, the simple counterfactual account of causation does not yield the verdicts on causation needed for NCA to give the right verdicts about harming. The problem, I take it, is that the account reinvites counterfactual comparisons to an allegedly non-comparative account of harming. If we want to keep NCA truly non-comparative, we need to couple it with some non-comparative analysis of causation.

Accounts that build on the idea of minimal sufficiency seems a better fit for NCA. An elementary version of such accounts can be stated as follows:

> *Elementary minimal sufficiency: C* causes *E* if and only if
> (i) *C* belongs to a set of actual antecedent events that guarantees, given the relevant laws, that *E* will occur, and
> (ii) if you remove *C* from the set, the set no longer guarantees that *E* will occur.[11]

Given this account of causation, NCA gives the right verdict in the cases we have considered so far. To begin with, it entails that Björn's and Dan's recommendations harmed me. Take Björn's recommendation for instance. It belonged to a set of actual antecedent events that included his recommendation but not Dan's. This set guaranteed, given the rules of the journal, that I would be sad and have few chances of getting tenure, and it did so regardless of whether Dan recommended rejection.

---

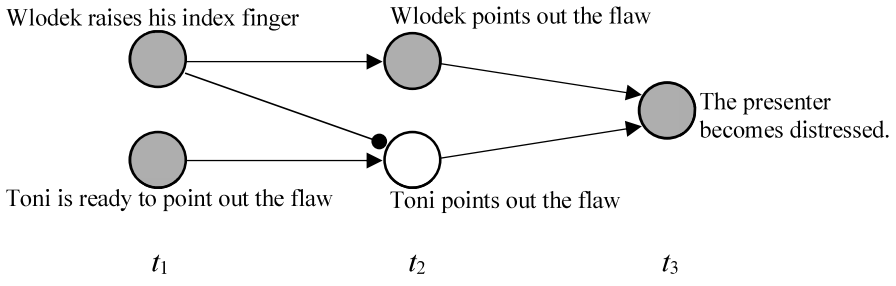
[10] Lewis (1973, 2004) presents more elaborated versions of this account.

[11] Mackie's (1974) INUS condition for causation and Wright's (1985) NESS condition for causation provides examples of such accounts.

Still, the set had not guaranteed this outcome if we had removed Björn's recommendation from it. Björn's recommendation is necessary for the sufficiency of that set. Therefore, Björn's recommendation caused me to be in an intrinsically bad state (sad, and with few chances of getting tenure), and hence NCA entails that Björn's recommendation harmed me. A parallel argument shows that Dan's recommendation harmed me.

However, *Elementary minimal sufficiency* needs to be elaborated to give the right verdict in *Presentation on pre-empted harm.* As this account is formulated now, it entails that both Wlodek and Toni caused the presenter's becoming distressed. At the time when Wlodek raised his index finger, Toni's readiness to point out the crucial flaw in the presenter's argument was minimally sufficient for the presenter's becoming distressed. It belonged to a set (not including Wlodek raising his index finger) of actual events that guaranteed, given the relevant laws, that the presenter would become distressed, and if we removed Toni's readiness from this set, it would no longer so guarantee. Still, it is obviously false that Toni caused the presenter's distress. As things turned out, Toni's readiness to point out the flaw in the argument was pre-empted before by Wlodek raising his index finger. It only guaranteed the outcome, but did not cause it.

The situation can be illustrated using a neuron diagram, where circles (or "neurons") represent events, and arrows represent causal connections. A shaded circle indicates that the neuron fires; that is, that the event occurs. The temporal reading is from left to right. If a neuron fires, it sends a signal through the connecting arrow to the right. If a neuron receives such a signal from its left, it will fire. A line ending in a black dot represents an inhibitory signal, hindering the neuron at its end point from firing.



As the diagram makes clear, events relevant for the causal evaluation of the situation occur in the intermediate time after Toni was ready to point out the flaw but before the presenter becomes distressed. At t_2, Toni decides not to raise his finger when he sees that Wlodek raises his. However, *Elementary minimal sufficiency* only takes into account what happens at time t_1 and t_3 when evaluating whether Toni's

readiness caused the presenter's distress. Following Caroline Touborg (2018), we can elaborate a more accurate account of minimal sufficiency that lets us capture what happens also at intermediate times, as follows:

Let us say that there is an *apparent process* from C to E when there is a chain of relations of minimal sufficiency. That is, when C belongs to a set of simultaneous events that is minimally sufficient a later event for D_1, D_1 belongs to a set of simultaneous events that is minimally sufficient for a later event D_2, … , and D_n belongs to a set of simultaneous events that is minimally sufficient for E. When we consider the line of events more closely, and consider more intermediate events between C and E, we might sometimes find that the apparent process was not genuine. That is, we might find intermediate times when the chain is broken. In such cases, we should conclude that C is not a cause of E. Conversely, if the chain is not broken when we consider more and more intermediate times, there is a genuine process connecting C to E, and we should conclude that C causes E. Call this account of causation *Elaborated minimal sufficiency.*

In *Presentation on pre-empted harm,* we find that the apparent process is not genuine. When we bring t_2 into the consideration, we find that the chain of relations of minimal sufficiency connecting Toni's readiness to point out the flaw to the presenter's becoming distressed is broken. As evidenced by the fact that Toni did not point out the flaw even though he was ready to do so, Toni's readiness to point out the flaw does not belong to a set of actually occurring events at t_1 that guaranteed the occurrence of this intermediate event at t_2. Thus, *Elaborated minimal sufficiency* correctly entails that Toni did not cause the presenter to become distressed, and so we can use it together with NCA to show that Toni did not harm the presenter.

Still, *Elaborated minimal sufficiency* entails that Wlodek raising his index finger was a cause of the presenter's becoming distressed (via his pointing out the flaw), so together with NCA it correctly entails that Wlodek harmed the presenter.

Have we found an accurate account of harming? Unfortunately, we have not. Still, our exploration of NCA and minimal sufficiency has not been in vain. We can use *Elaborated minimal sufficiency* together with SCA to construct an accurate account of harming. For what remains of this paper, I will first go through a few reasons why NCA coupled with *Elaborated minimal sufficiency* fails as an account of harming, and then go on to suggest a more accurate account.

Why NCA Fails

NCA together with *Elaborated minimal sufficiency* fails as an account of harming for a number of reasons. First, it gives counterintuitive verdicts in so called switching cases, such as the following:

> *The assistant's choice:* The assistant of the editor of the *Journal of Over-Determination* has the task of notifying me that my paper has been rejected. He can do this in two ways: either he can send the message himself by email, or he can wait until the journal's application system sends me the notification automatically. Either way, I will get notified, whereby I will get sad. As it turns out, the assistant sends the notification himself by email.

Here, it seems that the assistant does not harm me by sending the message via email.[12] I would get the message anyway, and there was nothing he could do to avoid me getting sad. However, this is not what *Elaborated minimal sufficiency* coupled with NCA entails. Remember that we have to consider intermediate events between the point in time when the assistant sends the email and the point in time when I receive the message. At one such point in time, the email arrives at my computer. We then find that the assistant's sending the message was minimally sufficient for the message's arriving at my computer, and that the message's arriving at my computer was minimally sufficient for me reading it and getting sad. So *Elaborated minimal sufficiency* entails that the assistant caused me getting sad. There is a genuine process going from the assistant's sending the email and my getting sad. NCA, in turn, then counterintuitively entails that the assistant harmed me.

Second, NCA combined with *Elaborated minimal sufficiency* cannot explain our shifting intuitions in *Journal of Over-Determination.* Remember that it seems on the one hand that Björn did not harm me since, given that Dan recommended rejection, I would become sad and have few chances of getting tenure regardless of whether Björn recommended rejection. Similarly, given Björn's recommendation to reject, it seems that Dan's recommendation did not harm me. On the other hand, it seems that Björn's and Dan's recommendations harmed me since, if they had not recommended rejection, I would not have become sad and I would have better chances of getting tenure. However, NCA combined with *Elaborated minimal sufficiency* entails that Björn's recommendation did harm me. It belongs to a set of simultaneous events that was minimally sufficient for me getting sad and with few chances of getting tenure. And the chain of relations of minimal sufficiency remains when we consider more and more intermediate times. The same goes for Dan's recommendation. So, rather than explaining our shifting intuitions, the account of harm under consideration brutely entails that one of the intuitions is correct.

Third, as Michael Rabenberg (2014), Molly Gardner (2017) and others point out, NCA gives counterintuitive verdicts in some cases regardless of which account of causation we use. Consider for instance the following case:

---

[12] Switching cases are common in literature on causation and on moral responsibility. See e.g. Thomson (1976) and Paul and Hall (2013). Standardly, the agent in such cases is not considered to cause the outcome, to be blameworthy for the outcome or to have control over the outcome. In a similar vein, it seems to me that what the agent in does in switching cases does not stand in right relation to the person *S* suffering harm for it to be correct to say that what the agent does harms *S*. In switching cases, unlike in pre-emption cases, there is no relevant possibility that the harm does not occur.

> *Dan's phone call*: Björn is really sad because he and his best friend Dan are not neighbours anymore when Dan calls him on the phone just to say "hi". After the call, Björn is still sad, but much less so.

Here, it does not seem that Dan's call harmed Björn. On the contrary, Dan's call cheered him up. However, this is not what NCA entails. Assuming that Dan's call caused Björn to be less sad (which any plausible account of causation would imply), NCA entails that Dan's call harmed Björn. Being a bit sad is an intrinsically bad state (we might assume), so Dan's call caused Björn to be in an intrinsically bad state.

The first two objections indicate that *Elaborated minimal sufficiency* is an inadequate account of causation. The third objection indicates that NCA's insistence on taking intrinsically bad states into account is problematic. My suggestion is that we go back to considering comparative accounts of harming like SCA to avoid the last problem, and that we use *Elaborated minimal sufficiency* to correct for the problems SCA runs into.

Harming

SCA seems to give a necessary but not sufficient condition for harming. It gives intuitively correct verdicts in some cases, like *Journal of Over-Determination*. However, in other cases, like *Presentation on pre-empted harm*, there seems to be something missing. While SCA correctly entails that Wlodek harmed the presenter (in a way), it fails to pinpoint the reason why Toni did not.

Elaborated minimal sufficiency, in turn, seems to capture a necessary but not sufficient condition for causation. It gives the intuitively correct verdict that Wlodek but not Toni caused the presenter to become distressed in *Presentation on pre-empted harm*, but it fails to pinpoint a reason why the assistant's email was not a cause of my being sad in *The assistant's choice.* It also fails to explain why it seems that, given that Dan recommended to reject my paper, Björn's recommendation to reject my paper does not seem to be a cause of my getting and having few chances of getting tenure.

This suggests that we might combine the conditions given in SCA and *Elaborated minimal sufficiency* to get a more accurate account of harming with two necessary conditions, as follows:

> *Harming:* An event *C* harms a person *S* in H if and only if
> (a) There is a genuine process connecting *C* to *S*'s being worse off, and
> (b) *S*'s being better off would have been more secure in H had *C* not occurred.[13]

[13] This account of harming is inspired by Touborg's (2018) account of causation and is in many aspects similar to the accounts of teleological reasons and blameworthiness Touborg and I have developed elsewhere.

This account gives the right verdict in all the cases we have considered. To begin with, it can explain our shifting intuitions in *Journal of Over-Determination.* Consider first the larger possibility horizon H_J according to which there are four relevant possible worlds: Björn and Dan recommends rejection, Björn but not Dan recommends rejection, Dan but not Björn recommends rejection, and neither Björn nor Dan recommends rejection. Given this possibility horizon, *Harming* entails that both Björn and Dan individually harmed me (as opposed to, for instance, merely being *involved* in harming me). As we have already seen, (a) there is a genuine process connecting Björn's recommendation to my being sad and having few chances of getting tenure, and (b) my not being sad and having few chances of getting tenure would be more secure in H_J if Björn had not recommended rejection. A similar argument shows that Dan's recommendation to reject my paper harmed me, given H_J.

Further, given the smaller possibility horizon $H_{J\text{-small}}$, where we do not treat it as a relevant possibility that Dan had recommended otherwise, *Harming* instead entails that Björn's recommendation to reject my paper did not harm me. Even though there still is a genuine process connecting his recommendation to my being sad and having few chances of getting tenure, his recommendation does not make this outcome more secure. In fact, there is no relevant possibility in this possibility horizon that I would have been better off. That is, while condition (a) still is satisfied, condition (b) is not. A parallel argument shows that Dan's recommendation did not harm me, given an alternative $H_{J\text{-small}}$ that holds fixed Björn's recommendation to reject.

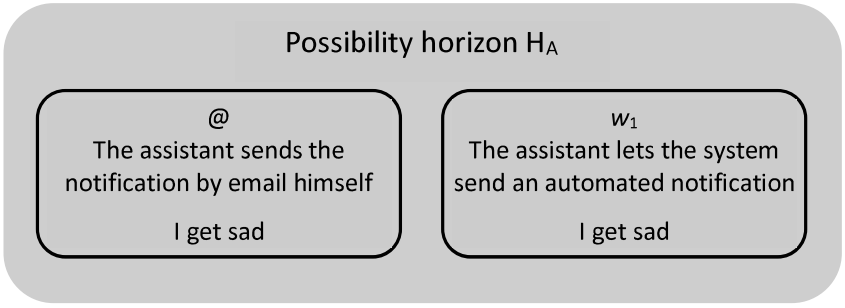
This way, *Harming* allows us to explain the shifting intuitions as stemming from different possibility horizons, or from different ways of understanding the situation. This raises the further question of which way to understand the situation that is the more accurate one. *Harming* does not help us out here. It does not say anything about which possibility horizon we should use. However, there are pragmatic reasons for thinking that the larger possibility horizon is the more accurate one. For one thing, it seems arbitrary to say that while there is a relevant possibility that Björn would have recommended otherwise, there is no relevant possibility that Dan would have done so (and vice versa). If we treat it as on open possibility that one of them could have recommended otherwise, there is a pressure to accept that both of them could have recommended otherwise.[14]

Harming also gives the right verdict that Wlodek but not Toni harmed the presenter (in a way) in *Presentation on pre-empted harm.* As we have seen, (b) both Wlodek raising his finger and Toni's readiness to point out the crucial flaw increased the security of the presenter's becoming distressed in H_P. This is why SCA wrongly indicates that Toni caused the presenter to become distressed. However, (a) there is a genuine process connecting Wlodek raising his finger (via his getting

[14] For more arguments why the larger possibility horizon typically is the more accurate one, see e.g. Gunnemyr (2021: 237-42).

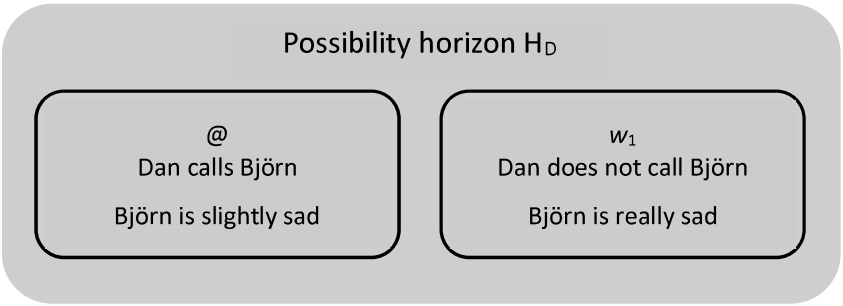
the word from the chair and pointing out the crucial flaw in the presenter's argument) to the presenter's becoming distressed, but no similar process connecting Toni's readiness to point out the flaw in the argument to this outcome. Therefore, *Harming* entails that Wlodek but not Toni harmed the presenter.

Next, *Harming* correctly entails that the assistant did not harm me by sending me an email himself rather than letting the system send me an automated notification in *The assistant's choice*. As explained earlier, (a) his sending the email is connected to my being sad via a genuine process. (This was why *Elaborated minimal sufficiency* gave the wrong verdict about the case.) Still, (b) his sending the notification by email did not increase the security of my sadness. To see this, we have to settle the relevant possibility horizon. In this case, there are two relevant possibilities, as indicated in the following possibility horizon:



There is no relevant possible world where I am better off; I get sad in every relevant possible world. Therefore, the security of the outcome that I get sad is just as secure in all relevant possible worlds. So, the assistant's sending the email does not make it less secure that I am better off, which means that condition (b) is not satisfied.

In addition, *Harming* yields the intuitively correct verdict that Dan did not harm Björn in *Dan's phone call.* While there is a genuine process connecting Dan's call to Björn's being (only) slightly sad, Björn's being better off had not been more secure if Dan had not called him. On the contrary, Björn would have been worse off had Dan not called him. To see this clearly, consider the relevant possibility horizon:



As you see, Björn is better off in @ than in w_1. Moreover, Björn's being better off would not have been more secure had Dan not called him. Indeed, had Dan not called him, Björn would have been worse off. He would have been really sad rather than just slightly sad.

Conclusions and Further Questions

To sum up, *Harming* gives intuitively correct verdicts about when an event harms someone in a wide range of cases. It gives the right verdict in overdetermination cases like *Journal of Over-Determination,* in pre-emption cases like *Presentation on pre-empted harm,* switching cases like *The assistant's choice,* and cases like *Dan's phone call* where someone's harm is relieved but not fully so. *Harming* can also explain the shifting intuitions we might have in cases like *Journal of Over-determination*: the different intuitions stem from different possibility horizons – or less formally, from different ways of understanding the situation at hand.

There is still work to do. As it stands, *Harming* will deliver counterintuitive verdicts in late pre-emption cases, in cases where it matters which possible world we take to be the relevant contrast, and in non-threshold cases (i.e. collective harm cases without a threshold). To get an idea of how *Harming* could be modified to cover such cases as well, see Gunnemyr (2021: chs 5, 6, 11 and 12) To work out the details will have to be work for another day.[15]

References

Bradley, Ben (2009) *Well-being and death*. Oxford: Clarendon Press.

Bradley, Ben (2012) "Doing away with harm". *Philosophy and Phenomenological Research, 85*(2): 390-412.

Feinberg, Joel (1984) *Harm to others*. New York: Oxford University Press.

Feit, Neil (2015) "Plural harm". *Philosophy and Phenomenological Research, 90*(2): 361-88.

Feit, Neil (2016) "Comparative harm, creation and death". *Utilitas, 28*(2): 136-63.

Gardner, Molly (2017) "On the strength of the reason against harming". *Journal of Moral Philosophy, 14*(1): 73-87.

Gunnemyr, Mattias (2021) *Reasons, blame, and collective harms* (PhD thesis). Lund University.

[15] I want to thank Andrés G. Garcia, Jens Johansson, and Jakob Werkmäster for constructive input on an earlier version of the paper.

Harman, Elizabeth (2009) "Harming as causing harm" in M. A. Roberts & D. T. Wasserman (Eds.) *Harming future persons* (137-54). Dordrecht: Springer Netherlands.

Jackson, Frank (1997) "Which effects" in J. Dancy (Ed.), *Reading Parfit* (42-53). Oxford: Blackwell.

Johansson, Jens (2023) "Petersson on plural harm" in A. G. Garcia, M. Gunnemyr, & J. Werkmäster (Eds.) *Value, morality & social reality: Essays dedicated to dan egonsson, björn petersson & toni rønnow-rasmussen.* Lund: Department of Philosophy.

Johansson, Jens, & Olle Risberg (2019) "The preemption problem". *Philosophical Studies, 176*(2): 351-65.

Klocksiem, Justin (2012) "A defense of the counterfactual comparative account of harm". *American Philosophical Quarterly, 49*(4): 285-300.

Lewis, David (1973) "Causation". *The Journal of Philosophy, 70*(17): 556-67.

Lewis, David (2004) "Causation as influence (extended)" in J. Collins, N. Hall, & L. A. Paul (Eds.) *Causation and counterfactuals* (75–106). Cambridge, Mass.: MIT Press.

Mackie, John L. (1974) *The cement of the universe: A study of causation*. Oxford: Clarendon.

Norcross, Alastair (2005) "Harming in context". *Philosophical Studies, 123*(1/2): 149-73.

Parfit, Derek (1984) *Reasons and persons*. Oxford: Clarendon Press.

Paul, L. A., & Edward J. Hall (2013) *Causation: A user's guide*. Oxford: Oxford University Press.

Petersson, Björn (2004) "The second mistake in moral mathematics is not about the worth of mere participation". *Utilitas, 16*(03): 288-315.

Petersson, Björn (2018) "Over-determined harms and harmless pluralities". *Ethical Theory and Moral Practice, 21*(4): 841-50.

Rabenberg, Michael (2014) "Harm". *Journal of Ethics and Social Philosophy, 8*(3): 1-32.

Shiffrin, Seana Valentine (1999) "Wrongful life, procreative responsibility, and the significance of harm". *Legal Theory, 5*(2): 117-48.

Thomson, Judith Jarvis (1976) "Killing, letting die, and the trolley problem". *The Monist, 59*(2): 204-17.

Touborg, Caroline Torpe (2018) *The dual nature of causation: Two necessary and jointly sufficient conditions* (Doctoral thesis). University of St Andrews.

Wright, Richard W. (1985) "Causation in tort law". *California Law Review, 73*(6): 1735-828.