Received: 24 November 2022 | Accepted: 17 July 2023

DOI: 10.1111/theo.12486

DISCUSSION

# Unruh's hybrid account of harm

**Erik Carlson** | **Jens Johansson** | **Olle Risberg**

Uppsala University, Uppsala, Sweden

**Correspondence**
Jens Johansson, Uppsala University, Uppsala, Sweden.
Email: jens.johansson@filosofi.uu.se

**Funding information**
Riksbankens Jubileumsfond, Grant/Award Number: P21-0462; Vetenskapsrådet, Grant/Award Numbers: 2018-01361, 2020-01955

**Abstract**
Charlotte Unruh has recently put forward a hybrid account of what it is to suffer harm – one that combines comparative and non-comparative elements. We raise two problems for Unruh's account. The first concerns killing and death; the second concerns the causing of temporarily low or high welfare.

KEYWORDS
benefit, harm, hybrid account of harm

## 1 | INTRODUCTION

What is it to suffer harm? At least three types of views about this issue can be distinguished. According to *comparative* accounts, suffering harm essentially involves having a welfare level that is lower than some relevant baseline. For instance, on a *temporal* comparative account, the baseline is determined by how well off the subject was in the past:

> **The Temporal Comparative Account**: A person *S* suffers harm if and only if *S* is worse off than before.

According to *non-comparative* accounts, by contrast, what it is to suffer harm can be understood without reference to any baseline. One non-comparative view focuses on suffering *ill-being* or negative welfare (that is, that which is intrinsically bad for one):

> **The Ill-Being Account**: A person *S* suffers harm if and only if *S* suffers ill-being.

The third kind of view combines comparative and non-comparative elements. Charlotte Unruh (2022) has recently proposed such a *hybrid* view of harm. While there are many possible hybrid views, Unruh's preferred approach combines elements of the two views just formulated. Since she (like most others in the debate) also takes benefit to be analogous to harm, her account can be formulated as follows:

> **Unruh's Hybrid Account**: A person *S* suffers harm (enjoys benefit) if and only if either (i) *S* suffers ill-being (enjoys well-being), or (ii) *S*'s well-being (ill-being) is lower than before.

Unruh uses 'well-being' to mean 'positive welfare'; thus someone's well-being, in her sense, consists in those things that are intrinsically good for them.[1]

Unruh's main argument for her account is that, unlike a variety of 'pure' comparative and non-comparative accounts, such as the Temporal Comparative Account and the Ill-Being Account, it is extensionally adequate. Consider the following two cases (2022, p. 5):

> *Bad Start in Life*: Patient was born with a painful condition, and so has never been better off.

> *IQ*: A professor who has a very high IQ takes a drug that reduces her IQ by a few points.

According to Unruh, both Patient and the professor intuitively suffer harm. But the verdict that Patient suffers harm conflicts with the Temporal Comparative Account (given that Patient has never been better off), and the verdict that the professor suffers harm conflicts with the Ill-Being Account (given that the professor's losing a few IQ points need not involve anything *intrinsically* bad for her). Unruh's account, by contrast, accommodates both verdicts.

An immediate worry about Unruh's view is that it fails to tell a unified story about what it is to suffer harm. In response, Unruh suggests that her view allows that harm has a 'unified core' (2022, p. 7), as suffering ill-being and having lower well-being than before are both ways in which one's welfare can be *adversely affected*. Accordingly, despite being disjunctive, the account may nonetheless be able to explain what all and only cases of suffering harm have in common.

Relatedly, Unruh suggests that the *severity* of harm is proportional to the extent to which one's welfare is adversely affected. While Unruh leaves many details unspecified, she suggests that the severity of a temporal harm increases not only with 'the magnitude of the loss of well-being, but also [with] its duration' (2022, p. 11), and might depend on other factors as well.

Unruh's main motivation for theorising about suffering harm seems to be that this has implications for when an action harms a person. In particular, she holds that *harming* requires *harm*:

> [I]t seems plausible that any account of harming needs to presuppose an account of harm … Intuitively, for a behaviour (such as Ann's throwing the stone) to count as harming, the behaviour needs to be related, in an appropriate way, to an outcome that counts as a harm (such as Bob's broken nose). (2022, p. 2; see also p. 8)

Thus, she suggests, a correct theory about suffering harm can inform our judgements about when actions are harmful, and thereby guide our applications of moral prohibitions on harming (2022, p. 2).

Unruh's account is a highly interesting contribution to the debate on harm and harming. It is arguably the most developed hybrid account in the literature, and it is one of the most developed accounts that take suffering harm to be more fundamental than harming.[2] Nevertheless, in this article we shall point out what we take to be two sets of serious problems for Unruh's account. The first concerns killing and death (Section 2); the second concerns the causing of

[1]Further, Unruh defines well-being and ill-being 'with reference to [a person's] neutral welfare level' (2022, p. 6). In a sense, this makes also condition (i) in her account a comparative condition, with the neutral welfare level as the baseline. However, nothing of substance hangs on whether this condition is classified as comparative or non-comparative.

[2]See Hanser (2008) for another account of the latter sort.

temporarily low or high welfare (Section 3). Section 4 provides a brief diagnosis regarding the root of these problems.

## 2 | KILLING AND DEATH

> *Assassination*: Victim is a flourishing young person who has spent her entire life at the same positive welfare level and never suffered ill-being. Assassin gives Victim a drug that kills her painlessly in her sleep. If Assassin had not given her the drug, Victim would have gone on to have a long, happy future.

Assuming, as is standard, that death is not intrinsically bad for the person who dies – and more generally, that nothing that is simultaneous with or subsequent to a person's death is intrinsically bad for her – Victim does not suffer ill-being. If we also assume that a person occupies no welfare level (not even a neutral level) after her death, Victim is never worse off than before. Unruh's account thus implies that Victim does not suffer harm. Given her claim that the harmfulness of an action presupposes harm, it follows that Assassin's action does not harm Victim. This is highly counterintuitive. Moreover, if the harmfulness of an action presupposes harm, then the same presumably holds for events that are not actions, including the event of death. Given this, Unruh's account problematically implies that the event of death in no case harms the person who dies.[3]

Moreover, the claim that Victim does not suffer harm is hard to reconcile with Unruh's suggestion that the unified core of harm is that the person's welfare is adversely affected (Section 1). It seems clear that Assassin's killing Victim, as well as the event of Victim's death, adversely affects Victim's welfare (in particular, by depriving her of well-being). In that case, of course, Victim's welfare is adversely affected.[4]

It is not an option to respond to these problems by insisting that Victim suffers ill-being while she is dead (for example, on the grounds that being dead is itself intrinsically bad for her). As Unruh points out (2022, p. 2), an account of harm should be axiologically neutral: it should not presuppose the truth or falsity of any substantive theory of well-being. Clearly, any account of harm that relies on the claim that Victim suffers ill-being while dead fails to satisfy this condition. For that claim conflicts with many substantive theories of well-being, including the most prominent ones. For instance, it is incompatible with hedonism, as Victim experiences no pain while dead.

A response that might seem more promising is to reject the assumption that a dead person occupies no welfare level, and instead claim that dead people have neutral welfare. Since Victim's welfare level was positive before Assassin killed her, it then follows that her welfare level, when she is dead, is lower than before. Given posthumous neutral welfare, then, Unruh's account implies that Victim suffers harm, and thereby allows that Assassin's action and the event of death harm Victim.[5] While there are metaphysical worries about ascribing welfare levels to people at times at which they do not exist, dealing with those worries might be considered preferable to biting the bullet and accepting that Assassin's action and Victim's death are harmless.

[3]For parallel reasons, Unruh's account also implies that death and killing never *benefits* the person who dies, even if (say) the person is otherwise bound to experience intense pain and suffering. This result is also counterintuitive. Further, note that none of these problems arise for views which do not understand harming in terms of harm. For example, according to the *counterfactual comparative account*, which is currently the most popular view of harming, an event harms (benefits) a person just in case she would have been better (worse) off if that event had not occurred. This account straightforwardly yields that both Assassin's action and Victim's death harm Victim. See further Section 4.

[4]Analogous remarks apply to Unruh's *Lottery Ticket* case: 'Ann prevents Bob from buying a lottery ticket. If Bob had bought the ticket, he would have won' (2022, p. 8). Unruh's account implies that Bob does not suffer harm. Unruh argues that this implication, along with its further implication that Ann's action does not harm Bob, is plausible. However, it seems clear that Ann's action adversely affects Bob's welfare – in which case, of course, Bob's welfare is adversely affected.

[5]We assume that having neutral welfare and no well-being entails having zero well-being.

However, invoking posthumous neutral welfare gives rise to two serious problems for Unruh's account. First, given posthumous neutral welfare, any person whose welfare level before death was positive has, for all eternity after death, a lower welfare level than before death. Hence, Unruh's account now implies that any person whose welfare level before death was positive continues to suffer harm for all eternity. This is absurd. Even more absurdly, given Unruh's idea that the severity of harm increases with the duration of harm (see Section 1), it also follows that the harm gets increasingly more severe for all eternity. One might try to avoid these results by claiming that a dead person has a (neutral) welfare level only until the time at which the person would have died, had they not died when they actually did. But invoking counterfactual considerations is, of course, highly alien to Unruh's non-counterfactual account of harm.

Second, the appeal to posthumous neutral welfare also threatens to undermine the motivation for Unruh's account. Presumably, if people have neutral welfare after death, they also have neutral welfare before birth.[6] And this in turn appears to render the non-comparative part of Unruh's account superfluous. As already mentioned, Unruh includes this part in her account in order to handle cases such as *Bad Start in Life* – the kind of case that she takes to undermine the Temporal Comparative Account (Section 1). But given prenatal neutral welfare, Patient has in fact been better off before. More generally, whenever a person suffers ill-being, she is (in that respect) worse off than she was before her birth, if her welfare level before birth was neutral. Thus, given posthumous and prenatal neutral welfare, the Temporal Comparative Account implies that Victim suffers harm in *Assassination*, and also gives the result that Unruh takes to be correct in *Bad Start in Life*. Given posthumous and prenatal neutral welfare, then, there seems to be no reason to abandon the Temporal Comparative Account in favour of Unruh's more complex account.

Thus, whether or not one believes in posthumous welfare, cases involving killing and death pose serious problems for Unruh's account.

## 3 | TEMPORARILY LOW OR HIGH WELFARE

> *Welfare Boost*: Zero is bound to spend his entire life, starting at time $t_1$ and ending at time $t_{100}$, at a neutral welfare level, never receiving anything that is intrinsically good or bad for him. There is only one exception: if Zero were to take a pill at $t_{10}$, this would cause him to experience intense pleasure at $t_{11}$ (yielding a welfare level of 1000 at $t_{11}$), after which he would for the rest of his life take some mild pleasure in the memory of this experience (yielding a welfare level of 1 at each time from $t_{12}$ to $t_{100}$). Taking the pill would not affect his life in any other way. Zero does take the pill at $t_{10}$.

Intuitively, given Zero's circumstances, there is nothing negative to say about his taking the pill and what this action brings about. Unruh's account, however, fails to respect this judgement in a number of ways.

First, since Zero is much worse off for a long time after $t_{11}$ than he was at $t_{11}$, Unruh's account apparently has the result that he suffers severe temporal harm. While Zero is also much better off at $t_{11}$ than before, this temporal benefit has a much shorter duration. By contrast, there is only a small difference in magnitude between his loss of well-being during $t_{12}$–$t_{100}$ (999 units) and his gain in well-being at $t_{11}$ (1000 units). Furthermore, the extent to which Zero is better off at each time from $t_{12}$ to $t_{100}$ than he is before $t_{11}$ is also small (1 unit). Unruh's account thus seems to yield the result that Zero overall suffers temporal harm to a much greater

[6]For instance, given hedonism, there seems to be no difference in the welfare-relevant properties that one has before and after existing, since each period of non-existence involves neither pleasure nor pain. And a person's welfare level surely supervenes on her welfare-relevant properties. Similar points apply to other theories of welfare.

extent than he enjoys temporal benefit.[7] Again, temporal harm and benefit is the only kind of comparative harm and benefit that Unruh's account allows; in particular, counterfactual factors, such as the fact that Zero is better off than he would have been had he not taken the pill, are irrelevant. Unruh's account therefore seemingly implies that Zero overall suffers comparative harm to a much greater extent than he enjoys comparative benefit (whereas he would have undergone neither comparative harm nor comparative benefit had he not taken the pill). This implication is very unappealing.

Second, it seems clear that Zero's act of taking the pill does not harm him (not even pro tanto). However, it seems that Unruh has to say that it does. Her account entails that Zero is in a state that counts as a harm from $t_{12}$ to $t_{100}$. While Unruh does not specify how, on her view, harmful actions have to be related to harmful states (see Section 1), any plausible candidate for that relation seems to obtain between the relevant action and the relevant state in *Welfare Boost*. For instance, Zero's taking the pill plausibly *causes* the state he is in from $t_{12}$ to $t_{100}$. Similarly, that state *counterfactually depends* on his taking the pill – had he not taken the pill, he would have had a welfare level of 0, instead of 1, from $t_{12}$ to $t_{100}$. Moreover, Zero's action is part of what *explains* his being in that state from $t_{12}$ to $t_{100}$. And so forth.[8]

Third, we take it to be clear that in the choice between taking the pill and refraining from taking it, Zero has no prudential reason at all (not even a weak pro tanto reason) to refrain. But it seems that Zero would have such a reason if Unruh's account were true. The following principle is highly plausible:

> **Prudential Reason**: Let *a* and *a** be two alternative actions available to *S* in a given choice situation. If *a* but not *a** would result in a state that is a harm for *S*, then *S* has a prudential reason to do *a** rather than *a*.

Again, Unruh's account implies that taking the pill results in a harm for Zero – the state that he is in from $t_{12}$ to $t_{100}$. Her account also implies that refraining from taking the pill would not result in a harm for Zero; he would never suffer any comparative or non-comparative harm. By Prudential Reason, then, in the choice between taking the pill and refraining from taking it, Zero has a prudential reason to do the latter rather than the former. This implication is unacceptable. Moreover, provided that the strength of the reason corresponds to the degree of the relevant harm, it also follows, even more implausibly, that Zero has a *strong* reason to refrain from taking the pill.

Arguably, a case similar to *Welfare Boost*, but involving temporarily *low* welfare, is even more problematic for Unruh's account:

> *Welfare Dip*: Zilch is bound to spend his entire life, starting at time $t_1$ and ending at time $t_{100}$, at a neutral welfare level, never receiving anything that is intrinsically good or bad for him. There is only one exception: if Zilch were to take a pill at $t_{10}$, this would cause him to experience intense pain at $t_{11}$ (yielding a welfare level of

[7] As we saw in Section 1, Unruh does not rule out that factors other than the magnitude and duration of the loss of well-being might affect the severity of temporal harm. But it is hard to see what further factors she could plausibly appeal to that are congenial to her account, in order to avoid the implication at hand.

[8] An anonymous referee has suggested that Unruh might respond by appealing to something like the following principle: An action harms *S* by resulting in *S*'s suffering a temporal harm only if *S* has a lower well-being level at the time of the temporal harm than *at or before the time of action*. Since Zero's well-being level from $t_{12}$ to $t_{100}$ is not lower than his well-being level at or before $t_{10}$, this principle allows Unruh to deny that Zero's taking the pill harms him by resulting in his suffering a temporal harm (and, of course, to deny that it harms him in any other way). It seems to us, however, that Unruh should not appeal to this principle. Suppose that Ten has a neutral well-being level at every time until and including $t_1$, and that Ten's well-being level at $t_2$ is bound to be 10 independently of what he does. Moreover, Ten's well-being level at $t_3$ will also be 10 unless he takes a pill at $t_1$. Ten does take the pill at $t_1$, as a result of which his well-being level at $t_3$ is only 1. Clearly, Unruh would want to say that Ten's taking the pill at $t_1$ harms him by resulting in his suffering a temporal harm at $t_3$. But the proposed principle implies that it does not, as Ten does not have a lower well-being level at $t_3$ than at or before $t_1$.

> −1000 at $t_{11}$), after which he would for the rest of his life have slightly unpleasant memories of this experience (yielding a welfare level of −1 at each time from $t_{12}$ to $t_{100}$). Taking the pill would not affect his life in any other way. Zilch does take the pill at $t_{10}$.

For reasons analogous to those we noted concerning *Welfare Boost*, Unruh's account seems to imply that Zilch enjoys comparative benefit to a much greater extent than he suffers comparative harm, that his action of taking the pill benefits him (at least pro tanto), and that he has a prudential reason to take the pill rather than refrain. These conclusions appear even more implausible than the parallel conclusions in *Welfare Boost*. Surely, there is nothing positive to say about Zilch's taking the pill or the consequences of this action.

## 4 | CONCLUDING REMARKS

The problems we have raised for Unruh's account concern both undergeneration and overgeneration. In *Assassination*, her account undergenerates, by leading to the implausible conclusion that neither Assassin's killing Victim nor Victim's death harm Victim. And in *Welfare Boost*, her account overgenerates, by leading to the implausible conclusions that Zero suffers severe comparative harm after taking the welfare-boosting pill, that his act of taking that pill harms him, and, by Prudential Reason, that he has a prudential reason not to take the pill. *Welfare Dip* illustrates that her account faces similar overgeneration problems concerning benefit.

One reason why some of these problems arise, we think, is that Unruh's account does not take seriously enough the plausible idea that an event's harming a person is a matter of its adversely affecting her welfare.[9] (For a defence of that idea, see Johansson & Risberg 2022.) As we pointed out in Section 1, in *Assassination*, both Assassin's killing Victim and Victim's death clearly affect Victim's welfare adversely. And in *Welfare Boost*, Zero's taking the pill clearly does not adversely affect his welfare. Yet Unruh's account implies, given the close connection that she takes to hold between harming and harm, that the former events are harmless whereas the latter event is harmful.

A possible response to some of our objections is thus for Unruh to abandon the close connection between harm and harming. On the resulting approach, an account of harm need not tell us anything at all about harming. However, while this would let her avoid the problematic verdicts about harming that we have emphasised, it would also risk depriving her account of much of its significance. As already mentioned (Section 1), Unruh's main motivation for investigating the nature of harm seems to be that this can yield a better understanding of when an action harms someone, and thereby help us determine when various moral prohibitions on harming apply. If her account of harm does not tell us anything about harming, this motivation is undermined. Of course, actions are not the only entities whose harmfulness is worthy of attention. For instance, the question of whether the event of death harms the person who dies is philosophically interesting and widely discussed, and not only because of its relevance to moral issues. However, since this question, too, concerns harming, an account of harm that does not tell us anything about harming will not help us answer it.

It would be costly for Unruh, then, to abandon her idea that harm is closely related to harming. This is not to suggest that we agree with her about the relationship between harm and harming; in particular, her suggestion that an account of harming needs to presuppose an account of harm (see Section 1) seems questionable to us. Compare: if one wants to understand

[9]Relatedly, recall that in Section 2 we argued that Unruh's account fails to accommodate her own idea that a common core of *harm* is the person's welfare being adversely affected.

what it is to kick someone, one would presumably not start by theorising about the notion of a 'state of being kicked', on the purported ground that any account of kicking needs to presuppose an account of that state. A better idea is to focus directly on the verbal notion; that is, on what it is to kick someone. Similarly, much of the debate about harming has focused directly on the relevant verbal notion – that is, on what it is for an event (such as an action) to harm someone – and it seems to us that Unruh does not give a sufficient reason to change this focus. In particular, in cases involving killing and death, it is at best only in some deflated sense that a harmful event (for example, Assassin's killing Victim) leaves the dead person in a 'harmed state'. This, we suggest, is another reason why Unruh's account goes wrong.

## ACKNOWLEDGEMENTS

Thanks to two anonymous referees for very helpful comments. Erik Carlson's and Jens Johansson's work on this paper was supported by Grant 2018-01361 from Vetenskapsrådet and Grant P21-0462 from Riksbankens Jubileumsfond. Olle Risberg's work on this paper was supported by Grant 2020-01955 from Vetenskapsrådet.

## ORCID

*Jens Johansson* https://orcid.org/0000-0002-2484-4502

## REFERENCES

Hanser, M. (2008) The metaphysics of harm. *Philosophy and Phenomenological Research*, 77(2), 421–450.

Johansson, J. & Risberg, O. (2022) A simple analysis of harm. *Ergo*, 9(19), 509–536.

Unruh, C. (2022) A hybrid account of harm. *Australasian Journal of Philosophy*, 1–14. Published online ahead of print. Available from: https://doi.org/10.1080/00048402.2022.2048401

**How to cite this article:** Carlson, E., Johansson, J. & Risberg, O. (2023) Unruh's hybrid account of harm. *Theoria*, 89(5), 748–754. Available from: https://doi.org/10.1111/theo.12486