



Harming and failing to benefit: a reply to purves

Jens Johansson¹  · Olle Risberg¹

Published online: 8 March 2019
© The Author(s) 2019

Abstract A prominent objection to the counterfactual comparative account of harm is that it classifies as harmful some events that are, intuitively, mere failures to benefit. In an attempt to solve this problem, Duncan Purves has recently proposed a novel version of the counterfactual comparative account, which relies on a distinction between *making* upshots happen and *allowing* upshots to happen. In this response, we argue that Purves’s account is unsuccessful. It fails in cases where an action makes the subject occupy a high well-being level though one of the available alternatives would have made it even higher. In fact, it fails even in some cases where each of the available alternatives to the action that was actually performed would have made the subject’s well-being level lower.

Keywords Harm · Benefit · Counterfactual comparative account of harm

1 Introduction

In a recent article in this journal, Duncan Purves discusses a common objection to one prominent view of the nature of harm: the counterfactual comparative account, or CCA (Purves 2018). According to CCA, a harmful event is simply one that leaves an individual worse off than he or she would have been in its absence—or as Purves formulates it (following Klocksiem 2012):

✉ Jens Johansson
jens.johansson@filosofi.uu.se
Olle Risberg
olle.risberg@filosofi.uu.se

¹ Department of Philosophy, Uppsala University, Box 627, 75126 Uppsala, Sweden

CCA An event e is a harm for a subject S just in case S is better off in the nearest possible world in which e does not occur than S is in the relevant e -world.¹ (Purves 2018, p. 2)

The objection is that CCA cannot distinguish harm from failure to benefit; more precisely, CCA appears to overgenerate harm by classifying as harmful some events that are, intuitively, mere (that is, not harmful) failures to benefit. Here is one of Purves's examples (building on a case in Bradley 2012):

Robin's Clubs⁺ Batman purchases golf clubs with the intention of giving them to Robin. As Batman is loading the clubs into the trunk of the Batmobile, it occurs to him that he hasn't checked the clubs for defects. Batman takes a few swings with the clubs. Because the feel of the clubs is much better than his own clubs, Batman decides to keep them. He tucks them away in his garage. (Purves 2018, p. 6)

Assuming that Batman would have given Robin the golf clubs had he not swung them, and Robin would have been better off with the clubs than without them, CCA implies that Batman's swinging the clubs harms Robin. Intuitively, however, this is a clear case of mere failure to benefit.

Purves notes that the problem also has a moral dimension. Common sense recognizes a moral asymmetry between harm and benefit, in that "harming someone is a much graver moral offense than failing to benefit him" (Purves 2018, p. 4). It is difficult for proponents of CCA to accommodate this asymmetry. To illustrate this difficulty, Purves contrasts Robin's Clubs⁺ with the following case (taken from Feit 2017):

Joker's Removal Batman has delivered golf clubs to Robin, and the Joker removes the clubs just before Robin opens the door to retrieve them. Had the Joker not removed the clubs from Robin's porch, Robin would have found them and thereby been better off. (Purves 2018, p. 4)

Intuitively, the Joker's action is morally wrong, whereas Batman's action in Robin's Clubs⁺ is morally permissible. A straightforward and commonsensical explanation of this moral difference is that the Joker's removing the clubs is a harm for Robin, whereas Batman's swinging the clubs is not. However, this explanation is unavailable for advocates of CCA, which counts both actions as harms.

In short, the failure to benefit problem apparently shows CCA to be both extensionally inadequate (classifying some nonharms as harms), and, as a result of this, at odds with ordinary morality. Some writers have responded that CCA's implications, in the relevant range of cases, are in the end acceptable: apparent instances of mere failure to benefit, which satisfy the criterion in CCA, are on closer inspection harmful (Feit 2015, 2017; Hanna 2016; Klockslem 2012). Our initial

¹ As Purves stresses, what CCA, like its rivals, aims to capture is *overall, all things considered* harm—overall harm as opposed to *pro tanto* harm, or harm in some respect; all things considered harm as opposed to merely intrinsic or merely extrinsic harm (Purves 2018, p. 2).

reaction to a case like Robin's Clubs⁺, these writers argue, is thus incorrect. Purves's approach is different. His aim is to develop a modified version of CCA that preserves our initial judgment that Batman's action is not harmful (and thereby much easier to justify morally than the Joker's). By emphasizing the difference between an event's *making* the subject and its merely *allowing* the subject to occupy his or her well-being level, Purves's account promises to "distinguish, in an intuitively satisfying way, between harms and mere failures to benefit" (Purves 2018, p. 5).

In this response, we shall argue that Purves's proposal, despite its ingenuity, is unsuccessful. His account fails in cases where the relevant action makes the subject occupy a high well-being level though one of the agent's alternative actions would have made it even higher (Sect. 3). Indeed, it fails even in some cases where each of the agent's available alternatives to the relevant action would have made the subject's well-being level lower (Sect. 4).

2 Purves's account

Purves argues that CCA's main competitors—in particular, temporal comparative accounts and noncomparative accounts of harm—need a distinction between making upshots happen and merely allowing upshots to happen. In order to deal with the failure to benefit problem, Purves suggests, CCA should be revised so as to include this element as well. He formulates the resulting account, "Harming as Making" (HAM), as follows:

HAM An event *e* is a harm (benefit) for S if and only if (1) *e* makes S occupy S's well-being level in the *e*-world, and (2) S's well-being level is higher (lower) in the nearest world in which *e* does not occur.² (Purves 2018, p. 15)

According to Purves, HAM has intuitively correct results in the central cases. In Robin's Clubs⁺, although Robin would have had a higher well-being level were it not for Batman's swinging the clubs, Batman's action does not plausibly *make* Robin occupy the well-being level that he in fact has. By contrast, Purves says, in Joker's Removal, the Joker's taking the clubs from Robin's porch does not only satisfy the counterfactual condition but also makes Robin occupy his well-being level. In other words, whereas the Joker's action satisfies both (1) and (2), Batman's action satisfies only (2) and not (1). On HAM, then, Batman's action does not harm Robin, whereas the Joker's action does—which is just the result that Purves wanted to achieve.

Presumably, most of us already have a reasonably good grasp of the distinction between making and allowing. In one of Purves's cases, you roll the bowling ball down the lane, while your bowling partner simply decides not to trip you during your approach. Without consulting any theory, it is natural to judge that your action makes the pins fall, whereas your partner's decision merely allows them to fall.

² HAM also includes a principle about the conditions under which an event is a mere failure to benefit (harm), which is less relevant to our current purposes. Thus we have omitted it.

Likewise, many will probably find intuitively correct Purves's claim that the Joker's removing the clubs, unlike Batman's swinging the clubs, makes Robin occupy his well-being level. However, Purves does not solely appeal to intuition. Drawing heavily on Fiona Woollard's account (Woollard 2008, 2015), he provides the following general criterion for making an upshot happen:

- (i) An event e makes U occur if and only if e is part of the sequence leading to U , where e is part of the sequence leading to U if and only if a fact about e is relevant to U through a complete chain of substantial facts.³ (Purves 2018, p. 15)

Purves gives elaborate definitions of various components of this account, rendering his position as a whole a rather complex one. While there is no need to go into the details here, Purves carefully argues that the Joker's removing the clubs is part of the sequence leading to Robin's occupying his well-being level, whereas Batman's swinging the clubs is not. Arguably, then, Purves's judgments about Batman's and the Joker's actions, with regard to making versus merely allowing, have both intuition and theory on their side.

For present purposes, we should note one condition that, according to Purves, is sufficient (though not necessary) for a fact's being "substantial" in the sense relevant to (i). A fact is substantial, Purves says, if it is "specificity positive"—that is, "a fact about what *is* the case rather than a fact about what is not the case" (Purves 2018, p. 13; emphasis his).⁴ The fact that you roll the bowling ball down the lane, for example, is relevant to the falling of the pins through a complete series of specificity positive facts. Hence your action makes the pins fall. By contrast, the fact that your partner decides not to trip you is relevant to the same upshot through at least some facts which are not specificity positive—such as, for example, the fact that your partner did *not* trip you. Nor is the fact that Batman swings the clubs, or the fact that the Joker removes the clubs, relevant to Robin's occupying his well-being level through a complete chain of specificity positive facts [though as indicated above, Purves's view has the result that the facts involved in the Joker's case are substantial for other reasons (see Purves 2018, pp. 18–19)]. We mention specificity positivity here since this phenomenon seems to provide the simplest kind of example of making rather than merely allowing. Presumably, the most straightforward way for an event to make an upshot happen is for some fact about the event to be relevant to the upshot through a complete chain of specificity

³ Purves also gives a criterion for merely allowing an upshot to happen. As with the corresponding element in HAM (see footnote 2), we have omitted it, as it is less relevant to our purposes.

⁴ What does this, in turn, amount to? Purves mentions the following partial explication (taken from Bennett 1995; via Woollard 2015): "a fact about A's behavior is specificity positive, just in case most of the ways A could have moved would have failed to make the corresponding proposition about A's behavior true" (Purves 2018, p. 13, fn. 32).

positive facts. Unfortunately for Purves, however, nothing more intricate than this simple phenomenon is needed to undermine HAM.

3 More golf

When a person fails to benefit someone, there are two different things that might be happening. One possibility is that the person fails to perform *any* action which benefits the subject. Robin's Clubs⁺ illustrates this possibility—Batman not only fails to give Robin the clubs, but he also does nothing else which benefits Robin. Another possibility, however, is that a person fails to benefit a subject by failing to perform some *specific* action which would have benefited the subject if it were performed. This possibility, unlike the first one, is intuitively compatible with the person performing *another* action which benefits the subject. HAM might perhaps have plausible results with regard to many cases of the first type, but with regard to the second type of case, its implications are often highly counterintuitive. Here is one example:

More Golf Batman has bought two sets of golf clubs—one that is of extremely high quality and one that is somewhat worse but still very good. Now he has three options: (A1) to give Robin the slightly worse golf clubs; (A2) to give Robin the better golf clubs; and (A3) to keep all the golf clubs for himself. Batman performs (A1), which makes Robin's well-being level increase significantly. If Batman had not done so, he would have given Robin the even better set of golf clubs, whereby Robin's well-being level would have been even higher.

In this case, Batman fails to perform one action that would have benefited Robin; namely, the action of giving him the best set of golf clubs. Intuitively, however, this is not a case in which Batman does *nothing* that benefits Robin. On the contrary, the action which Batman does perform—namely, the action of giving Robin the set of golf clubs that is merely very good—also appears to be a benefit to Robin. At the very least, Batman's action of giving Robin the somewhat worse set of golf clubs cannot plausibly be said to be a *harm* to Robin—that view is not only implausible in its own right but also one that we should be especially eager to reject if we think, with Purves, that in Robin's Clubs⁺, Batman's swinging the golf clubs does not harm Robin. The problem is that this view about the harmfulness of Batman's action is precisely what HAM entails. For Batman's giving Robin the somewhat worse set of golf clubs satisfies both conditions in the theory: (1) it makes Robin occupy the well-being level that he in fact occupies (as, we may naturally suppose, it is relevant to that outcome through a series of specificity positive facts, like the fact that Robin was happy to receive the gift, that he used the clubs in future golf sessions, and so on); and (2) had it not been performed, Robin would have been (even) better off. HAM thus implies, wrongly, that Batman's giving Robin the somewhat worse set of golf clubs harms Robin. Hence HAM is false.

To be clear, the problem is not that HAM treats Batman's respective actions in Robin's Clubs⁺ and More Golf asymmetrically, in saying that one but not both of them harms Robin. Such an asymmetrical treatment might perhaps be defensible. If it is, however, then surely the action that is harmful is not Batman's giving the somewhat worse set of golf clubs to Robin in More Golf, but his swinging the clubs in Robin's Clubs⁺. After all, the former but not the latter action makes Robin occupy a very high well-being level even though Batman could easily have acted so that Robin would have occupied a much lower well-being level. Insofar as an asymmetrical treatment is defensible, then, it is the reverse of what HAM yields.

As with CCA, furthermore, the fact that HAM appears to be extensionally inadequate also makes it difficult to reconcile with common sense views about the moral relevance of harm. Intuitively, insofar as Batman acts morally permissibly in Robin's Clubs⁺, it is even more plausible that he does so in More Golf. A straightforward explanation of the moral permissibility of these two actions is that both are harmless. But that explanation is unavailable to proponents of HAM, as it entails that what Batman does in More Golf is a harm to Robin.

Finally, given some further natural assumptions about More Golf, it also illustrates that HAM is incompatible with the following Very Plausible Principle about harm:

VPP Let A and B be two actions in the same alternative set. If (i) A would not harm a subject S if it were performed, and (ii) B, if it were performed, would make S have a well-being level that is both very high and much higher than S's well-being level would have been if A were performed, then (iii) B would not harm S if it were performed.

VPP is a more specific version of the intuition that in a given situation, performing an action that harms a person cannot be much better for them than it would have been to perform an alternative action that would be harmless to them. That HAM is incompatible with this principle can be shown as follows. We have already stipulated that More Golf is such that relative to the nearest (A1)-world—which happens to be the actual world—the nearest \sim (A1)-world is an (A2)-world.⁵ In other words, had (A1) been performed (as indeed it was), then, had it not been performed, (A2) would have been performed instead. Let us now also stipulate that relative to the nearest (A3)-world—that is, the nearest world in which Batman keeps both sets of golf clubs for himself—the nearest \sim (A3)-world is an (A2)-world. What this means is that (A3) is such that if it had been performed, then, if it had not been performed, (A2) would have been performed instead.

Now, HAM entails that if (A3) were performed, it would not harm Robin. The reason is that (A3) is in all relevant respects comparable to Batman's action in Robin's Clubs⁺, and thus, on Purves's view, (A3) would not *make* Robin have the well-being level that he would have. However, if (A1) were performed—that is, if Batman gave Robin the somewhat worse set of golf clubs—then that action would make Robin have a well-being level that is both very high and much higher than

⁵ An (A1)-world is a world in which (A1) is performed; an \sim (A1)-world is a world in which (A1) is not performed; and so forth.

Robin's well-being level would have been if Batman had performed (A3). Yet as shown above, HAM also entails that (A1), unlike (A3), would be a harm to Robin if it were performed. Thus VPP and HAM are incompatible—and the view to be rejected, we submit, is the latter.

4 Darts

In *More Golf*, one of Batman's available alternatives to the action that he actually performed would have made Robin occupy an even higher well-being level. It is worth noting that HAM fails even in some scenarios where this is not the case—in some scenarios, that is, where the relevant action *maximizes* the subject's well-being. The problem is due to the fact that in a given situation, the relevant action might be such that if it had not been performed, then none of the agent's available alternatives would have been performed either.

To illustrate the problem, we can use a revised version of a case from a recent paper by Erik Carlson (who does not focus on the failure to benefit problem, and does not discuss Purves's account).

Darts Batman faces a choice between either throwing a dart at a circular board, which is surrounded by a much thinner circle of a different color, or throwing the dart outside of both the board and the surrounding circle. Robin will get \$100,000 if Batman hits the board, \$101,000 if Batman hits the surrounding circle, and nothing if Batman hits neither. The more money Robin has, the higher his well-being level. Batman is quite good at throwing darts, so he can hit either the board or the area outside of both the board and the thin surrounding circle. These are his only two alternatives, though—hitting the thin surrounding circle is not an action that is available to him. He is not extremely good at throwing darts, after all, and he would not succeed in hitting the thin circle even if he tried. What happens is that Batman hits the board. If he had not hit the board, he would have tried but failed to do so, and thereby hit the surrounding circle.⁶

⁶ Cf. Carlson (forthcoming: Sect. 2). Notably, a variant of the case above, that is more similar to Carlson's original case, also shows that CCA (see Sect. 1), like HAM, is incompatible with VPP (see Sect. 3). To see this, suppose that in addition to what we described above, there is also a second board that Batman can hit, which does not have a surrounding circle. If Batman were to hit the second board, Robin would get no more than \$15. Finally, suppose that in the nearest world where Batman hits the second board, it is true that were he not to do so, he would miss both the boards and the circle surrounding the first board.

Now, let A in VPP be Batman's hitting the second board and let B in VPP be Batman's hitting the first board. CCA entails that (i) A would not harm Robin if it were performed (since if A were performed, Robin would not have been better off if A were not performed). Further, it is true that (ii) B, if it were performed, would make Robin have a well-being level that is very high and much higher than it would be if A were performed. On VPP, it thus follows that (iii) B would not harm Robin if it were performed. But that verdict is in conflict with CCA, since it is true of B that were it performed, then, were it not performed, Robin would have been (even) better off.

Batman's hitting the board evidently *makes* Robin occupy his actual well-being level; in particular, there is no reason to doubt that it is relevant to Robin's occupying his well-being level through a complete series of specificity positive facts (such as the fact that Batman's hitting the board activates the relevant mechanism which transfers the money to Robin's bank account, and so on). Furthermore, if Batman had not hit the board, Robin would have occupied an even higher well-being level. In other words, Batman's hitting the board satisfies both conditions in HAM. Thus, HAM entails—as does CCA—that Batman's hitting the board harms Robin. As with More Golf, this implication is not merely implausible in its own right, but also one that it would be especially unreasonable to accept for anyone who, like Purves, wants to deny that Batman's swinging the clubs in Robin's Clubs⁺ harms Robin. After all, Batman's hitting the board, unlike his swinging the clubs, makes Robin occupy a high well-being level even though Batman could easily have acted so that Robin's well-being level would have been much lower. Already for these by now familiar sorts of reasons, then, Darts is a serious problem for HAM.

But there is a further reason as well. For unlike Robin's Clubs⁺ and More Golf, Darts is also a case where no alternative action available to Batman is such that, if it had been performed, Robin would have occupied a higher (or even equally high) well-being level. While Batman's hitting the surrounding circle is something that *could* have happened—and again, is something that *would* have happened if Batman had not hit the board—it is not an alternative action available to Batman.⁷ In other words, it is not something that Batman *can* do, in the sense of 'can' that 'ought' is commonly taken to imply. Hence, Batman's hitting the board not only makes Robin occupy a high well-being level, but also *maximizes* Robin's well-being level—and, indeed, it is the only action among his available alternatives to do so. It would not be sensible to say that Batman's hitting the board nevertheless harms Robin, and it would be even less sensible to add that, by contrast, Batman's swinging the clubs in Robin's Clubs⁺—which does *not* make Robin occupy a high well-being level, and does *not* maximize Robin's well-being level—does *not* harm Robin. Yet that is precisely what we would have to say if HAM were true.

Recall, finally, that Purves intends his view to accord with commonsense morality, especially the view that “harming someone is a much graver moral offense than failing to benefit him” (Purves 2018, p. 4). On HAM, as we have seen,

⁷ Why is it not an available action? In our view, a strong reason for thinking so is that hitting the surrounding circle is not under Batman's intentional control, as he would not succeed even if he tried, and there is also no natural sense in which Batman *knows how* to hit the surrounding circle. But it also seems to us that in this context, nothing really depends on the detail. Instead, whatever account of performability that one prefers, we take it that any plausible view about the matter must entail that hitting the surrounding circle is not an action that is available to Batman in this case.

An anonymous reviewer objected that a claim about performability along the following lines is at least prima facie plausible: If it could have happened that S did something, A, that would bring about some state *p*, and S knows that their doing A would bring about *p*, then A is an alternative available to S. But, while it is true that Batman's hitting the surrounding board plausibly satisfies both conditions of the antecedent, the claim as a whole is not plausible. Consider unintentional sneezing. This is something that we often could do, since we often do it. And we typically also know that we would bring about various states of the world by unintentionally sneezing (e.g., that we unintentionally sneeze). But unintentionally sneezing is normally not something that we can do in the sense of 'can' that 'ought' plausibly implies.

Batman's swinging the clubs in Robin's Clubs⁺ merely fails to benefit Robin, whereas his hitting the board in Darts harms Robin. Yet if one of these two actions is a graver moral offense than the other, surely it is the former rather than the latter. While Batman's swinging the clubs seems morally acceptable, there might be at least some weak moral reason against it, since, after all, one of the alternative actions would have made Robin better off. By contrast, there is no moral reason at all against Batman's action in Darts.

5 Conclusion

The counterfactual comparative account of harm is beset by the failure to benefit problem. While HAM, Purves's modified version of the counterfactual comparative account, is supposed to solve that problem, we have argued that it fails to do so. Instead, HAM too classifies as harms certain actions which are intuitively nonharms. For that reason, it also comes into conflict with common sense morality.

Of course, it remains to be seen whether there is any other version of the counterfactual approach to harm that avoids these problems. That is not a question that we have discussed in this paper. What we have sought to show is just that, if such a theory does exist, it probably does not lie in the direction of Purves's view.

Acknowledgements Thanks to Erik Carlson and an anonymous reviewer for helpful comments on this paper. Parts of Olle Risberg's work on this paper were generously supported by Helge Ax:son Johnson's Foundation, Hultengren's Foundation, The Royal Swedish Academy of Letters, History and Antiquities, The Salén Foundation, Sixten Gemzén's Foundation, Thun's Grant Foundation, and Värmland's nation in Uppsala. Jens Johansson's work on this paper was generously supported by Grant P14-0212:1 from Riksbankens Jubileumsfond.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bennett, J. (1995). *The act itself*. Oxford: Oxford University Press.
- Bradley, B. (2012). Doing away with harm. *Philosophy and Phenomenological Research*, 85, 390–412.
- Carlson, E. (forthcoming). More problems for the counterfactual comparative account of harm and benefit. *Ethical Theory and Moral Practice*.
- Feit, N. (2015). Plural harm. *Philosophy and Phenomenological Research*, 90, 361–388.
- Feit, N. (2017). Harming by failing to benefit. *Ethical Theory and Moral Practice*. <https://doi.org/10.1007/s10677-017-9838-6>.
- Hanna, N. (2016). Harm: omission, preemption, freedom. *Philosophy and Phenomenological Research*, 93, 251–273.
- Klocksien, J. (2012). A defense of the counterfactual comparative account of harm. *American Philosophical Quarterly*, 49, 285–300.
- Purves, D. (2018). Harming as making worse off. *Philosophical Studies*. <https://doi.org/10.1007/s11098-018-1144-1>.

- Woollard, F. (2008). Doing and allowing, threats and sequences. *Pacific Philosophical Quarterly*, 89, 261–277.
- Woollard, F. (2015). *Doing and allowing harm*. Oxford: Oxford University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.