

# The preemption problem

Jens Johansson<sup>1</sup>  · Olle Risberg<sup>1</sup>

Published online: 2 December 2017

© The Author(s) 2017. This article is an open access publication

**Abstract** According to the standard version of the counterfactual comparative account of harm, an event is overall harmful for an individual if and only if she would have been on balance better off if it had not occurred. This view faces the “preemption problem.” In the recent literature, there are various ingenious attempts to deal with this problem, some of which involve slight additions to, or modifications of, the counterfactual comparative account. We argue, however, that none of these attempts work, and that the preemption problem continues to haunt the counterfactual comparative account.

**Keywords** Harm · Preemption · Well-being

## 1 Introduction

The debate on harm is dominated by the counterfactual comparative account, the most straightforward version of which can be formulated as follows:

CCA An event  $e$  is overall harmful for  $S$  if and only if  $S$  would have been on balance better off if  $e$  had not occurred

CCA has many virtues beyond the obvious ones of elegance and simplicity. For one thing, it is in accord with the plausible idea that harming is a matter of making some sort of difference for the worse. For another, CCA has intuitively appealing implications in various important cases. In particular, while noncounterfactual

---

✉ Jens Johansson  
jens.johansson@filosofi.uu.se

Olle Risberg  
olle.risberg@filosofi.uu.se

<sup>1</sup> Department of Philosophy, Uppsala University, Box 627, 75126 Uppsala, Sweden

accounts have severe problems accounting for the harm of death (Purves 2016), CCA yields that death is overall harmful for someone whose continued life would have been happy, though not for someone whose continued life would have been miserable. Moreover, because CCA is merely concerned with *overall* harm, i.e., harmfulness all things considered, it is compatible with the reasonable claim that the latter person's death can be *pro tanto* harmful for her, i.e., harmful in some respect, so long as her continued life would have contained some happy moments. (For the most part, we shall keep the "overall" qualification silent.)

However, CCA faces some well-known problems. One common objection to CCA concerns *omission*. Ben Bradley provides the following case:

Suppose Batman purchases a set of golf clubs with the intention of giving them to Robin, which would have made Robin happy. Batman tells the Joker about his intentions. The Joker says to Batman, "why not keep them for yourself?" Batman is persuaded. He keeps his golf clubs. (Bradley 2012: 397)

According to Bradley, while Batman fails to benefit Robin, this does not constitute harming him. But if CCA were true, Bradley suggests, it would do so.

Neil Feit (Feit 2015) and Nathan Hanna (Hanna 2016) have recently argued, however, that the omission problem is less serious than it initially appears. While it would be implausible to say that someone is harmed whenever it is not the case that she is benefited, Feit and Hanna emphasize that views like CCA imply no such thing. In particular, unless we make the unattractive assumption that nonoccurrences are events, there are myriads of cases where an individual fails to be benefited but where no actual event is such that, had it not occurred, then she would have been better off. Contrary to what Bradley suggests, then, CCA does not imply that Batman's mere failure to benefit Robin constitutes harming him. It is thus easy to overestimate the extent to which, intuitively, CCA overgenerates harm. In order for CCA to have the result that Batman does something that harms Robin, he has to perform some action—for instance, make a decision—such that Robin would have been better off if it had not been performed. It is not evident that he does; as Hanna points out, even assuming that Batman makes a decision not to give Robin the clubs, it might be that if he had not made this decision, then he would have remained indecisive, or forgotten to decide (Hanna 2016: 252). But, Feit and Hanna argue, if the decision really does have the required feature, then it is after all defensible to say that it harms Robin (although Feit concedes that the problem still constitutes "a strike" against CCA (Feit 2015: 385)).<sup>1</sup>

Another objection concerns *preemption*. Here the problem is not, as in the omission case, that CCA ascribes harmfulness to apparently harmless events, but that it ascribes harmlessness to apparently harmful events. Nathan Hanna offers a concise presentation of a case from Alastair Norcross (Norcross 2005: 165–166; see also Bradley 2012: 407; Rabenberg 2015: 10):

<sup>1</sup> Hanna also gives an interesting argument for the claim that Batman's action harms Robin, based on the claim that the Joker's action harms Robin (Hanna 2016: 253–257).

Bobby Knight gets mad at a philosopher, Phil, in response to a perceived slight. So he chokes Phil. Luckily for Phil, Knight is in anger management. It's taught him to better control his behavior. He choked Phil because he applied anger management techniques. If he hadn't, he would have dismembered Phil. (Hanna 2016: 263)

Intuitively, Knight's act of choking Phil harms Phil; but CCA implies that it does not.

The preemption problem is not restricted to cases in which the pertinent events are actions. Here are Bradley—and Batman—again:

Suppose Batman drops dead of a heart attack. A millisecond after his death, his body is hit by a flaming cannonball. The cannonball would have killed Batman if he had still been alive. (Bradley 2012: 397)

CCA entails that the heart attack does not harm Batman, but intuitively, it does.

As more or less all participants in the debate agree, preemption cases pose a serious challenge to CCA. Some people like challenges, however, and so a number of ingenious attempts to deal with the preemption problem can be found in the recent literature. In Sect. 2 we discuss the proposals by David Boonin (Boonin 2014) and Justin Klockslem (Klockslem 2012), each of whom argues that CCA as it stands can handle the problem.<sup>2</sup> In Sect. 3 we turn to Feit's view, which entails CCA but also involves a crucial addition, the notion of *plural* harm: the harmfulness attributable not only to a single event, but to several events taken together. In Sect. 4 we discuss Hanna's approach. Although Hanna, unlike the others, believes that the preemption problem shows CCA to be mistaken in cases where the relevant event is a free action, he suggests that a modified version of CCA avoids the problem. We seek to show that all of these proposals—as well as various variants of them that will suggest themselves in the course of our discussion—are unsuccessful. While the reasons for this differ, one recurring theme will be that the proposed solutions fail in cases where one of the agent's available alternatives was to simply leave the other person alone. The preemption problem thus continues to haunt the counterfactual comparative account of harm; indeed, as we briefly indicate in our concluding remarks (Sect. 5), there may be reason to give alternative theories another chance.

## 2 Unnecessary evils

Because Boonin's discussion of the preemption problem occurs in the context of the "non-identity problem," which concerns the moral status of certain procreative actions, he focuses on cases where the pertinent event is an action. Most importantly for our purposes, Boonin claims that CCA's implication in the Bobby Knight case—

<sup>2</sup> To be exact, Boonin commits himself only to the component of CCA that states a necessary condition on harm (which is the component threatened by the preemption problem). He does not commit himself to the component that states a sufficient condition on harm, although he finds it attractive (Boonin 2014: 53).

that the choking does not harm Phil—is acceptable.<sup>3</sup> The reason for this, he suggests, is that the case is relevantly similar to other cases, which apparently pose no problem for CCA. Suppose, Boonin says, that the meter by your parked car (whose left door is left unlocked) has expired, and that just as the police officer

is about to write the \$25 ticket, he instead decides to give you a break by taking [four quarters] in change from your car to feed the meter on your behalf. If there is no way that the police officer is going to simply leave your car there without either ticketing it or feeding the meter with your quarters, then it seems ... that the act of taking the change from your car and putting it in the meter does not harm you. And it ... seems plausible to say this precisely because the alternative would have made things even worse for you. (Boonin 2014: 62)

According to Boonin, since this case is structurally symmetric to the Bobby Knight case, we should also deny that Knight's action harms Phil. If choking Phil “really is the only thing that will in fact prevent [Knight] from tearing [Phil's] arms off, then his [choking Phil] does not really harm [Phil]” (Boonin 2014: 62–63).

It is true that the two cases—the Bobby Knight case and ticket case—can be constructed so that the agent's only alternative to his actual act—choking Phil, and taking the change from your car and putting it in the meter, respectively—is an act that would make the other person even worse off. Then the actual act is a sort of necessary evil, the only available way for the agent to prevent something even more detrimental to the other person's well-being. Maybe on this assumption it is reasonable to say that the actual act is merely *pro tanto* harmful, but not overall harmful, for the other person. However, nothing forces us to construct the cases in this way. We are perfectly free to stipulate, for instance, that one of the alternatives available to Bobby Knight was to leave Phil alone—to avoid hurting him in any way at all. This stipulation is compatible with the supposition that leaving Phil alone is not what Knight *would* have done if he had not choked him (since he would then have dismembered him). For example, due to Knight's anger it might require slightly more effort on his part to leave Phil alone than to dismember him, although Knight was able to perform either alternative. Given the stipulation, it is very implausible to deny that the choking is overall harmful for Phil. After all, it caused him serious pain and discomfort although one of Knight's available alternatives would have left him entirely unscathed.<sup>4</sup>

<sup>3</sup> We shall write as if Boonin's comments concern the Bobby Knight case as presented in Sect. 1; his own version of the case differs only trivially from it (Boonin 2014: 62).

<sup>4</sup> Two further comments on this. First, Duncan Purves has suggested to us that if Bobby Knight knows that he will dismember Phil if he does not choke him, then it is not very implausible to deny that the choking overall harms Phil. It is not clear to us that such knowledge is compatible with the assumption that leaving Phil alone is one of Knight's alternatives. In any case, we can simply stipulate that Knight does not possess such knowledge.

Second, our stipulation that Knight has yet another available alternative is mainly meant to prime the reader's intuitions. We do not mean to suggest that a correct account of harm should make reference to such a factor. (Cf. the final two paragraphs of Sect. 4.) Thanks to an anonymous referee for prompting us to make this clarification.

Similarly, we are free to construct the ticket case so that one of the police officer's available alternatives was to leave your financial status entirely intact, by neither ticketing your car nor taking any of your quarters. This stipulation is compatible with the supposition that this is not what he *would* have done had he not acted as he did (for instance, because he thinks you would thereby get away a little too easily). Given the stipulation, it is no longer plausible to deny that his taking the quarters from your car and putting it in the meter harms you.<sup>5</sup> Maybe the police officer (unlike, presumably, Bobby Knight) acts permissibly, and maybe your negligence makes you deserve being harmed; but none of this means that his act does not harm you. Thus, when the ticket case is made relevantly similar to the modified Bobby Knight case, the analogy is of no help for the CCA proponent, but for the CCA opponent.

Justin Klocksiem's defense of CCA is vulnerable to basically the same problem. Counterfactuals, Klocksiem writes, "are highly sensitive to context, and there may be no single way things would have been if things had gone otherwise" (Klocksiem 2012: 289). A sentence of the form, "If *e* had not occurred, then *S* would have been better off" can be true or false depending on what features of the case are the most salient, something that will vary depending on what specific aspects we are interested in, which in turn varies from context to context. Given CCA's reliance on counterfactuals of this sort, then, harm judgments themselves are highly sensitive to context. According to Klocksiem, this observation helps us handle a host of problems for CCA, including the preemption problem. When the salient comparison world, among all possible worlds in which a given actual event does not occur, is one in which the person is better off, we are inclined to say that the event harms her; when the salient comparison world is one in which she is not better off, we are not so inclined. For instance, Klocksiem suggests, one agent's breaking your legs strikes us as something that overall harms you, provided that our context is such that the salient comparison world is one in which everyone leaves you alone, but not if the salient comparison world is one in which another agent is determined to kill you in case he finds you uninjured.

Even if this approach takes care of some problematic cases, it does not take care of the Bobby Knight case (which Klocksiem does not consider). After all, the relevant contextual details are already made explicit in the presentation of the case (see Sect. 1): it is explicitly stated that if Knight had not choked Phil, then he would have dismembered him. There is thus no question that, with regard to the counterfactual sentence, "If Knight had not choked Phil, then Phil would have been better off," the salient comparison world is one in which Knight dismembers Phil. Hence, our context is such that the sentence is false. However, this in no way prevents us from also assuming, without any relevant shift in context, that one of Knight's available alternatives was to leave Phil alone. Surely every sensible version of contextualism about counterfactuals must respect the distinction between an action that the agent *would* have performed instead of his actual action, and an

---

<sup>5</sup> Assuming, that is, that your well-being corresponds to your financial status in this case. Note, however, that is Boonin, and not us, who needs this assumption, as without it the original ticket case is not analogous to the original Bobby Knight case.

action that he *could* but *would* not have performed instead of his actual action. So long as we do assume that Knight was indeed able to Phil alone, it seems as clear as ever that the choking harms Phil, contrary to CCA.<sup>6</sup>

Similarly, consider Klockslem's leg-breaking case. It may well be true that in a context where the salient comparison world is one in which you are killed by another agent instead, the sentence "If the agent had not broken your legs, then you would have been better off" comes out false. However, we are still free to also stipulate, without any relevant shift in context, that one of the leg-breaker's available alternatives—for instance, spending all his effort on making the presumptive murderer calm down—would have left your legs and your life intact. On this assumption, it seems clear that his breaking your legs overall harms you, contrary to CCA.

### 3 Plural harm

The basic idea in Feit's "plural harm" approach is that some events can *together* harm a person—or, which is supposed to say the same thing, a *plurality* of events can harm a person—by making her worse off than she would have been if none of them had taken place. Feit offers the following, more precise principle:

PH E [overall] harms S if and only if E is the smallest plurality of events such that, if none of them had occurred, S would have been [on balance] better off. (Feit 2015: 371)

The "smallest plurality" restriction is designed to exclude superfluous events, that is, those that are not relevantly involved in depriving the person of well-being. Even if the plurality of events *e*<sub>1</sub>, *e*<sub>2</sub>, and Fred Astaire's dancing in the distance satisfies the condition that, had no event in it occurred, then an individual would have been better off, this plurality still does not harm her if the smaller plurality consisting only of *e*<sub>1</sub> and *e*<sub>2</sub> also satisfies this condition. We take it that "smallest plurality" in PH should be read as, "smallest plurality that includes some event in E"; surely PH is not meant to rule out that lots of nonoverlapping pluralities of many different sizes can harm an individual. Feit eventually modifies PH to deal with the problem that several pluralities can be tied for smallest, and to accommodate degrees of harm (Feit 2015: 376–79), but since those complications are irrelevant for present purposes, we will focus on PH.

PH is intended to also cover cases where a single event harms, or does not harm, someone. Feit assumes that a plurality can consist of only one event; more exactly, he makes it clear that just as speaking of a plurality of several events is just a way of speaking of *them*, speaking of a plurality of a single event is just a way of speaking of *it*. This means that PH entails CCA though not vice versa. Clearly, if a single event satisfies the condition that a person would have been better off if it had not occurred, then this event constitutes the smallest plurality, of those that include it, satisfying this condition.

<sup>6</sup> For a related criticism of Klockslem's treatment of the omission problem, see Purves (ms).

Before turning to PH's treatment of preemption cases, let us highlight a noteworthy implication of PH that Feit does not mention. Suppose an agent simultaneously performs two actions,  $e_1$  and  $e_2$ , which compose a further single compound action,  $e_3$  (not to be conflated with the plurality of  $e_1$  and  $e_2$ ). Suppose also that if neither  $e_1$  nor  $e_2$  had been performed, then you would have been better off than you actually are; however, if  $e_1$  but not  $e_2$  had been performed, then you would have been worse off than you actually are; and similarly, if  $e_2$  but not  $e_1$  had been performed, then you would have been worse off than you actually are. To illustrate,  $e_1$  may be the action of making you extremely painfully ill, while  $e_2$  is the action of force-feeding you a medicine which cures the disease with a negative side effect, such as leaving you seriously disabled. Suppose furthermore that if  $e_3$  had not been performed, then either  $e_1$  or  $e_2$  (though not both) would still have been performed. PH entails that the plurality of  $e_1$  and  $e_2$  harms you, although neither  $e_1$  nor  $e_2$  nor the action they compose,  $e_3$ , harms you; indeed, given a parallel account of benefit, each of those three acts benefits you. This is an interesting result. Is it a bad result? That is difficult to say. On the one hand, it does seem at least mildly paradoxical—particularly the claim that two actions together harm you whereas the action that they together make up benefits you. On the other hand, this sort of case might actually reveal an advantage of adopting PH over merely sticking to CCA. If CCA is true, and if there is no plural harm, then the agent does not do anything that harms you in this case; again, assuming a parallel account of benefit, everything she does benefits you. This is counterintuitive, and threatens to make the agent praiseworthy even though her actions have significantly lowered your well-being. On PH, by contrast, the agent does do something—or at least some things—that harm you, namely, the plurality of  $e_1$  and  $e_2$ .

In any case, PH's main merit is, naturally, supposed to be its capacity to deal with the preemption problem. On PH, Batman is harmed by the plurality of his heart attack and the cannonball's hitting his body, as he would have been better off had neither of these events occurred, although neither of them is such that Batman would have been better off had *it* not occurred. Of course, the latter fact means that PH implies—unsurprisingly, as PH implies CCA—that Batman's heart attack does not itself harm him. But this, Feit suggests, is no serious problem so long as we can say that Batman is harmed, or at least so long as we can say that the heart attack is one of the events that together harm him. Likewise, according to Feit it is “not an embarrassment” to have to say that in the Bobby Knight case, the choking itself is rendered harmless (and even beneficial): “It is similar to a case in which one person tortures another because this will prevent a more severe torture by a third person.” (Feit 2015: 381)

As we argued in the previous section, it *is* an embarrassment to render harmless Bobby Knight's choking Phil. Perhaps Feit is implicitly assuming that Knight had only two alternatives available to him: choking Phil or dismembering him. But again, we are free to stipulate that one of the alternatives available to Knight was to leave Phil alone, and then it is highly implausible to deny that the choking harms Phil. Similar remarks apply, moreover, to the case where torturing someone prevents her from being more severely tortured by a third person. Surely if one of the alternatives available to the agent was to do something that would have saved

the other person from being tortured at all—even if this is not what *would* have happened were it not for the actual torture—then the actual torture was overall harmful for her. Feit’s analogy strengthens rather than weakens the objection.

PH confronts another problem as well, one that Feit addresses but, in our view, seriously underestimates. In the heart attack case, the rest of the harmful plurality (the cannonball’s hitting Batman’s body) is easily found. Other cases, however, raise what Feit calls the problem of unavailable events. For instance, he notes that in the Bobby Knight case, he cannot say that the choking and the dismemberment together harm Phil, for the latter event is purely counterfactual: it never actually occurs (and thus fails to exist). However, Feit claims that the counterfactual that if Bobby Knight hadn’t choked Phil, then he would have dismembered him—as well as any true counterfactual involving preempting and preempted events in this way—must be grounded in something nonmodal. More specifically, for any true counterfactual of this sort, there must be some actual events “making [it] true” (Feit 2015: 280n.). In this particular case, Feit says, those actual events should be identified with “certain of Knight’s token psychological states (perhaps feelings of rage)” (Feit 2015: 381). Call these mental events the “M-events.” According to Feit, the M-events, together with the choking, are what harm Phil. Thus, whereas the choking itself does not harm Phil, it is one of the events that jointly do so.

This line of response is problematic for two reasons. First, it is implausible that the M-events make true the choking/dismemberment counterfactual (the claim that if Bobby Knight hadn’t choked Phil, then he would have dismembered him). While there is controversy over what exactly “making true” involves, an almost universally accepted view—which we take it that Feit does not wish to reject—is that it at least requires necessitation. That is, a thing makes a claim true only if, necessarily, if the thing exists then the claim is true (e.g., Merricks 2007: 5). The corresponding principle in the case of *joint* truthmaking (several events together making something true), which is what Feit’s suggestion concerns, is that some things jointly make a claim true only if, necessarily, if all of the things exist then the claim is true (Merricks 2007: 18). Applied to the case at hand, the requirement is thus that the M-events be such that, necessarily, if they exist, then the choking/dismemberment counterfactual is true. However, the M-events do not satisfy this requirement. For there are many possible worlds in which the M-events exist but in which the choking/dismemberment counterfactual is false: for instance, a possible world where the M-events exist and Phil is an expert in avoiding dismemberment. Moreover, it will not help much to simply expand the group of truthmaking events, that is, to say that the choking/dismemberment counterfactual is made true, not by the M-events all by themselves, but by the M-events together with various additional events. For it seems that there are even possible worlds in which *all* actual events exist but in which the choking/dismemberment counterfactual is false: for instance, a possible world where all actual events exist and where there is also a powerful demon who is determined to prevent anyone from dismembering Phil. Or is the nonoccurrence of the demon’s decision, or the nonexistence of the demon, itself an event (in which case the envisaged world is impossible, since some actual event is then absent from that world after all)? If so, then so is Batman’s, or



anyone's, not giving Robin the golf clubs. Hence, this move would undermine Feit's treatment of the omission problem (see Sect. 1).

Second, even granting that the M-events do make true the choking/dismemberment counterfactual, is Feit right to contend that on his view, the plurality consisting of the M-events and the choking harms Phil? There are two main ways of construing the case, and Feit's contention does not seem promising on either of them. On one construal of the case, the M-events are causally responsible for Knight's attacking Phil at all—or more exactly, if none of the M-events had occurred, then Knight would not have hurt Phil in any way. These feelings of rage, for instance, are thus not exclusively “dismemberment-oriented.” Perhaps the M-events do not cause Knight to choke Phil rather than to dismember him; that is more plausibly caused, or at least causally explained, by his applying anger management techniques. But the important point is that on the present construal of the case, if none of the M-events had occurred, then Knight would not have choked Phil. In that case, though, PH yields that Phil is harmed by the plurality consisting *solely* of the M-events (or perhaps, solely of some of them), that is, *without* the choking; for if no event in this smaller plurality had occurred, then Phil would have been better off. In other words, the plurality consisting of both the choking and the M-events is too big to harm Phil on PH.

Not only is it false here that the M-events and the choking together harm Phil on PH; there need not be *any* events that together with the choking harm Phil on PH. This is fatal for PH, for we take it to be unacceptable to simply embrace the claim that the choking, aside from not itself harming Phil, is not even one of the events that together harm Phil. As Feit seems to agree, that claim is deeply counterintuitive; the choking is nothing like Fred Astaire's dancing in the distance. Furthermore, being a consequence of some harmful events should not in this way deprive an event of its harming power.

Likewise, it would be a nonstarter to modify PH to no longer exclude superfluous events, so that any plurality of events counts as harming someone so long as she would have been better off if none of them had occurred. While the choking would then be one of the events that jointly harm Phil, it would be so in the same trivial way as, again, Fred Astaire's dancing in the distance.

On the other main way of construing of the case, if none of the M-events had occurred, then Bobby Knight would still have attacked Phil. Probably, then, whereas the M-events might be exclusively “dismemberment-oriented,” some of Bobby Knight's *other* token psychological states, the “M\*-events” (which are perhaps *other* feelings of rage), are causally responsible for his hurting Phil at all. Presumably, given this construal, if none of the M-events had occurred, the choking would still have occurred. This means that given this construal, the above problem is avoided: since getting rid of the M-events would not suffice to get rid of the choking, the plurality of the M-events and the choking is now not too big to harm Phil according to PH. This is cold comfort, however, for we still do not get the result that this plurality harms Phil on PH. Plausibly, if neither any of the M-events nor the choking had occurred, then the M\*-events would still have occurred. They would not have resulted in the choking, of course, but in light of the significant amount of hostility that they display toward Phil, there is little reason to suppose that they

would not have resulted in something equally detrimental to his well-being. After all, there are countless ways to hurt someone, and why believe that Knight would have pursued one of the softer ones, or none at all, if it were not for the choking (and the M-events)? Naturally, we can add the M\*-events to the supposedly harmful plurality—so that the resulting plurality consists of them as well as the M-events and the choking. It seems that if none of the events in this larger plurality had occurred, then Phil would have been better off. But this reintroduces the earlier problem; according to PH, this plurality does not harm Phil, as the choking is now superfluous.<sup>7</sup>

We do not deny that it is often highly natural to say that various events together harm an individual. But this is very limited evidence for Feit's view; indeed, reflection on some typical cases points away from it. Often when we say such things, what we want to emphasize is that none of the events would have harmed the individual in the absence of the other ones—even if each of the events is such that, if it had not occurred, then the individual would have been better off. For example, consider a non-preemption variant of the heart attack case. Suppose there are no cannonball events, but that Batman drops dead as a result of two independent and tiny heart attacks of equal strength, both of which are required to make him worse off; either of them would have made no difference to his well-being in the absence of the other one. These assumptions make it intuitive and natural to say that the two heart attacks together harm Batman. This is not to deny that each of them itself harms him (due to the presence of the other one); it is just a way of stressing that the harming is here a team effort. However, this is not the kind of plural harm that PH allows, as each of the heart attacks is such that, if it had not occurred, then Batman would have been better off. Thus, PH entails that the plurality of both of them does not harm Batman—that they do *not* together harm him.

## 4 Fundamental intentions

Hanna's version of the counterfactual comparative account is interestingly different from more orthodox views, as it entails that different principles hold for free actions and other events (such as heart attacks), respectively. More precisely, while Hanna is happy to accept CCA when restricted to events that are not free actions, he thinks

<sup>7</sup> Feit discusses an alternative to PH where "causal sequences of events," rather than single events, are "the basic units of harm" (2015: 383–384). That view does entail that the choking is one of the events which harm Phil. Roughly, the reason is that the choking is part of a causal sequence such that, if the sequence had not been initiated, Phil would have been better off. Feit does not endorse the appeal to causal sequences, however, and we think that he is right not to do so. One problem (partially noted by Feit with regard to a different case) is that this view entails that several events which occur after the choking—including, say, Phil's taking painkillers to relieve the pain—are comparable to the choking in terms of their harm to Phil; after all, those events too are part of a causal sequence such that, if the sequence had not been initiated, Phil would have been better off. This result is unacceptable. Surely, Knight's choking Phil is involved in harming him to a much greater extent than Phil's taking painkillers (an event that does not harm Phil at all, either by itself or together with other events).

that a different principle is required to account for the harmfulness of free actions.<sup>8</sup> Here is the one he proposes:

CCA+ An agent's free act is overall harmful for a subject S iff S is better off in the closest possible world where the agent doesn't freely perform an act of that type than S is as a result of the act. (Hanna 2016: 266)

The most important difference between CCA and CCA+ is the following: on CCA, whether a free action harms a subject depends on her well-being level in the nearest possible world in which the action is not performed; whereas on CCA+, it depends instead on her well-being level in the nearest possible world where an action of the relevant *type* is not performed. Thus, in cases where it is true that if a given free action had not been performed, then an action of the same type would have been performed in its place, CCA and CCA+ may have different implications.

It is this feature of the view that apparently allows Hanna to explain why Bobby Knight's act of choking Phil is harmful. Hanna claims that the choking and the dismemberment are, in the relevant sense, actions of the same type. On an ordinary account of action types, this claim would be difficult to maintain, since choking somebody and tearing their arms off are intuitively quite different types of actions. However, Hanna stipulates that for his purposes, we should say that two actions are of the same type just in case the agent performs them from the same *fundamental intention* (Hanna 2016: 267). And since Knight's fundamental intention is to hurt Phil, these two actions—choking Phil and dismembering Phil—are both ways for Knight to achieve what he fundamentally intends to achieve. Hence, on CCA+, whether the choking harms Phil does not depend on Phil's well-being level in the nearest world where Knight does not perform *that* action, but on Phil's well-being level in the nearest world where Knight does not perform *any* action with the intention of hurting Phil. That world is not a world in which Knight tears Phil's arms off; it is a world where Knight does not hurt him in any way at all. Since Phil's well-being level would have been higher if Knight had not hurt him at all, CCA+ (unlike CCA) entails that the choking harms Phil.

Hanna's nonstandard notion of an action type—on which, again, two actions are of the same type just in case the agent performed them from the same fundamental intention—obviously plays a central role in his account. He does not say much about how fundamental intentions, in turn, should be understood, but he does say that they have a "motivational priority" (Hanna 2016: 267) and must be cited in order to fully explain an agent's actual and counterfactual behavior. The idea appears to be that in order to explain why Knight choked Phil, and why Knight would have dismembered Phil if he had not choked him, we must cite Knight's fundamental intention of hurting Phil. Knight's ultimate aim, so to speak, is to hurt Phil, which is why he would have hurt Phil by dismemberment if he had not hurt

<sup>8</sup> Regarding preemption cases involving events that are not free actions, Hanna holds that the preempting event is *pro tanto* harmful but not overall harmful (Hanna 2016: 262–263; see also Kloockslem 2012). CCA is of course compatible with this result (at least given certain understandings of the notion of *pro tanto* harm and its relation to overall harm). In relation to such cases, Hanna also expresses some sympathy for Feit's account of plural harm, which we have already criticized (Hanna 2016: 261–262).

him by choking him. This is why the nearby dismemberment-world is irrelevant, on CCA+, for determining whether the choking harms Phil.

There remains room for dissatisfaction with the possible unclarity of Hanna's notion of a fundamental intention, and more generally, with the disjunctiveness of his overall account of harm. But these seem to be relatively minor issues, which we could live with if the account were otherwise plausible. We also grant that in the original version of the Bobby Knight case, CCA+ (unlike, as we have seen, Feit's view) does have the desired result that Knight's action harms Phil. However, we object that CCA+ has implausible results if we consider a wider range of preemption cases. In fact, we shall see that with regard to certain preemption cases, Hanna's view has *less* plausible results than CCA, over which CCA+ is intended to be an improvement. Since the main motivation for Hanna's view consists in its supposed ability to handle preemption cases involving free actions, this problem appears to be especially significant in this context: not only does it strongly suggest that the view is false, but it also undercuts the main reason for thinking that the view is true.

We shall present two preemption cases where CCA+ has intuitively incorrect results. The first case involves an apparently harmful act which preempts an act which would have made the subject even worse off, but which would have been done from a *different* fundamental intention. Consider this variant of the Bobby Knight case:

Bobby Knight gets mad at a philosopher, Phil, in response to a philosophical presentation which Phil gave in an annoying tone of voice. So even though Knight could simply have let Phil be, he chokes Phil with the fundamental intention of ensuring that nobody ever speaks with an annoying voice again. If Knight had not acted from that intention, he would have acted from the different intention of ensuring that Phil never writes a philosophy paper again. (The reason is that Knight recently read a paper by him which was very boring.) Thus, if Knight had acted from a different fundamental intention than the one from which he actually acted, he would have torn Phil's arms off instead.<sup>9</sup>

Intuitively, Phil is harmed by Knight's choking him in this case too. In that regard, this case does not seem relevantly different from the original version of the Bobby Knight case. But notice that in this case, Phil is worse off in nearest world where Knight performs an act which is of a different type than the act—choking Phil—that Knight actually performs. CCA+ thus implausibly entails that Phil is not harmed by

---

<sup>9</sup> We would like to make two comments about this case in relation to Hanna's discussion. First, since Knight could have freely chosen not to hurt Phil at all, this case is different from the case that Hanna calls "Dark Knight 2," in which it is assumed (roughly) that Knight can only avoid significantly lowering Phil's well-being level by slightly lowering his well-being level instead (Hanna 2016: 268–289). Second, we may, but need not, assume that Knight actually has the intention of ensuring that Phil never writes a philosophy paper but that he does not act on it. It is sufficient to assume that Knight *would* have acted on this intention if he had performed an act which is of a different type than the one which he actually performs. Thus, there need not be any sense in which Knight's action is "an act of restraint," if Hanna is correct to think that this issue matters for the overall assessment of the case (Hanna 2016: 267–268).

Knight's choking him. (CCA+ also has the even more implausible result that Phil is in fact benefited by Knight's choking him, if we assume, with Hanna (2016: 263), that benefit and harm should be treated analogously.)

We find it very plausible that the choking harms Phil in both the original Bobby Knight case and in the modified case just presented. Neither CCA nor CCA+ is compatible with that claim. However, this claim is also stronger than what is necessary to argue against CCA+: we only need the logically weaker and even more plausible claim that *if* Phil is harmed in the original Bobby Knight case, *then* he is also harmed in the modified version of the case. Even CCA is compatible with that assumption, as it entails that Knight is harmed in neither case. CCA+ is not, however, as it entails that Phil is harmed in the original Knight case but not in the modified case. Such a "mixed" view about the harm that is done in these cases is arguably even worse; it appears to accord way too much weight to whether Knight's preempted action would have been done from a different fundamental intention or not, which simply does not seem to matter for whether the choking harms Phil. CCA+ is, then, arguably worse than CCA with respect to this sort of preemption case.

Turn now to the second counterexample to CCA+. This one involves an action which is intuitively harmful since it causes the subject to occupy a low level of well-being and the subject would have been better off if the action had not been performed. However, if the agent had not acted from the fundamental intention from which he actually acts, he would have caused the subject to be at an equally low level of well-being anyway. (The action that he would then have performed is thus preempted by his acting from the fundamental intention from which he actually acts.) Let us again use Batman and the Joker as protagonists:

Batman is painfully shot by the Joker because the Joker intends to gain control over Gotham City. Had the Joker not performed this free act, he would have tried to befriend Batman instead, again with the fundamental intention of gaining control over Gotham City. The reason is that the Joker believes—perhaps falsely—that the second best means of gaining control over the city is to become friends with Batman. However, if the Joker had not acted from the intention of gaining control over the city at all, he would have acted from the different intention of ensuring that his reputation as a villain is nevertheless in good standing. In order to achieve that result, he would have shot Batman in a painful way, just as he does in the actual world.

Given these assumptions, Batman's well-being level is much higher in the nearest world in which the Joker does not shoot him. CCA thus entails that Batman is harmed by the Joker's shooting him in a painful way, and it seems to us that any plausible version of the counterfactual account should have this result. On CCA+, however, the question of what would have happened if the Joker had not shot Batman is irrelevant: the only question that matters on this view is what would have happened if the Joker had not performed an act of the relevant type at all. And in the nearest world in which the Joker does not act from the fundamental intention from which he actually acts, he shoots Batman, just like he actually does. Batman's well-being level in that world is thus precisely as low as it actually is. Accordingly,

CCA+ implausibly entails that Batman is not harmed by the Joker's shooting him. This is another sort of preemption case where the implications of CCA+ are less plausible than those of CCA.

In conclusion, we believe that Hanna has failed to solve the preemption problem. To end this section, we shall briefly consider the alternative, more simple view that an agent's free act harms a subject if and only if there is an alternative action which the agent could have performed instead such that, if that action had been performed, then the subject would have been better off.<sup>10</sup> This view may seem tempting in the present context, since it has intuitively plausible implications in both the original Bobby Knight case and the cases presented in this section. But this advantage comes at a great cost: even if this view avoids the preemption problem, it succumbs to an especially striking version of the omission problem instead—much more striking than the one faced by CCA and CCA+.

To see this, think about what you must do to make somebody as well off as you can. Barring exceptional circumstances, this will involve extreme sacrifices on your part, such as giving them everything you own, or traveling large distances to become their servant for life. Almost every second, there is a huge number of people whom you can benefit in some such way. Unlike CCA and CCA+, the present proposal entails that whatever you do instead, your action is harming all these people. Similarly, of course, there is a huge number of people who could benefit *you* in some similar ways, almost every second. Whenever they do something else instead, this view entails that you are harmed too. These results are surely too much to stomach. Even if we accept—as we should—that Bobby Knight's choking Phil is harmful when Knight could simply have left Phil alone, it does not follow that whenever Knight could have made Phil better off, what he actually does instead harms Phil.<sup>11</sup>

## 5 Concluding remarks

In this paper, we have discussed several ways of salvaging the counterfactual comparative theories of harm from the preemption problem. If our arguments have been sound, the preemption problem remains a serious difficulty for all these views. One may be tempted to add further modifications and extra clauses to the counterfactual account in order to try to avoid this objection, but we are skeptical of this enterprise. The preemption problem appears to reoccur in a systematic way for counterfactual theories of harm, even when they are modified with the precise aim of avoiding it. This suggests (though of course it does not entail) that the proper response is abandonment of the counterfactual account, and not further modification. For no matter how we specify which world we are to compare to the actual world, with regard to the relevant subject's level of well-being, it seems there will

---

<sup>10</sup> See Roberts (2007) for discussion of similar accounts.

<sup>11</sup> Thanks to Duncan Purves for asking us to clarify this.

always be cases where her well-being level in that world is entirely irrelevant for whether the subject is in fact harmed.

Thus, unless we should follow Bradley's advice and "[l]et harm go the way of phlogiston" (2012: 411), we are disposed to think that some of the many possible noncounterfactual theories of harm require further consideration. It is true that Bradley and others have convincingly argued that the most commonly discussed noncounterfactual theories have problems at least as serious as those on which we have focused here. But maybe other, significantly different noncounterfactual—or at least only partly counterfactual—views would fare better? In future harm research, we think, this area deserves our attention.

**Acknowledgements** We are grateful to Duncan Purves, an anonymous referee, and participants in the Higher Seminar in Practical Philosophy at Uppsala University for very helpful comments. Jens Johansson's work for this paper was 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

- Boonin, D. (2014). *The non-identity problem and the ethics of future people*. New York: Oxford University Press.
- Bradley, B. (2012). Doing away with harm. *Philosophy and Phenomenological Research*, 85, 390–412.
- Feit, N. (2015). Plural harm. *Philosophy and Phenomenological Research*, 90, 361–388.
- Hanna, N. (2016). Harm: Omission, preemption, freedom. *Philosophy and Phenomenological Research*, 93, 251–273.
- Klockslem, J. (2012). A defense of the counterfactual comparative account of harm. *American Philosophical Quarterly*, 49, 285–300.
- Merricks, T. (2007). *Truth and ontology*. Oxford: Clarendon Press.
- Norcross, A. (2005). Harming in context. *Philosophical Studies*, 123, 149–173.
- Purves, D. (2016). The harms of death. In M. Cholbi (Ed.), *Immortality and the philosophy of death* (pp. 67–84). Lanham, MD: Rowman & Littlefield.
- Purves, D. (ms). Harming as making worse off. Unpublished manuscript. Department of Philosophy, University of Florida.
- Rabenberg, M. (2015). Harm. *Journal of Ethics and Social Philosophy*, 8(3), 1–33.
- Roberts, M. (2007). The non-identity fallacy: Harm, probability and another look at Parfit's depletion example. *Utilitas*, 19, 267–311.