



# More Problems for the Counterfactual Comparative Account of Harm and Benefit

Erik Carlson<sup>1</sup>

Accepted: 17 September 2018 / Published online: 26 September 2018  
© The Author(s) 2018

## Abstract

The counterfactual comparative account of harm and benefit (CCA) has several virtues, but it also faces serious problems. I argue that CCA is incompatible with the prudential and moral relevance of harm and benefit. Some possible ways to revise or restrict CCA, in order to avoid this conclusion, are discussed and found wanting. Finally, I try to show that appealing to the context-sensitivity of counterfactuals, or to the alleged contrastive nature of harm and benefit, does not provide a solution.

**Keywords** Harm · Benefit · Counterfactual comparative account · Prudential reasons · Contextualism · Contrastivism

The counterfactual comparative account (CCA) is probably the most widely held account of harm and benefit. The aim of this paper is to show that CCA is incompatible with the prudential and moral relevance of harm and benefit. Since it seems implausible to deny that harm and benefit are prudentially and morally relevant, this is a serious problem for CCA. The paper is structured as follows. In section 1, I present CCA, and try to sort out some unclarities in the standard formulation of the account. In section 2, I present my main objections to CCA. In section 3, I discuss and reject some possible ways to revise or restrict CCA, so as to avoid my objections. In section 4, I try to show that appealing to the context-sensitivity of counterfactuals does not provide a solution. In section 5, I tentatively argue that reverting to a “contrastive” version of CCA is not of much help, either. Section 6 briefly concludes.

## 1 The Counterfactual Comparative Account

A standard formulation of CCA is as follows:

---

✉ Erik Carlson  
erik.carlson@filosofi.uu.se

<sup>1</sup> Department of philosophy, Uppsala University, Uppsala, Sweden

*The counterfactual comparative account of harm and benefit (CCA):* An event  $e$  or action  $a$  harms (benefits) a person overall if and only if she would have been on balance better (worse) off if  $e$  had not occurred, or  $a$  had not been performed.

Two clarificatory comments are in place. First, overall harm or benefit should be contrasted with pro tanto harm or benefit. If you save my life but break my leg in the process, you probably benefit me overall, although you do me pro tanto harm. In what follows, ‘harm’ and ‘benefit’ refer to overall harm and overall benefit, respectively. Also, ‘better off’, ‘worse off’, and related expressions should be read in the “on balance” sense.<sup>1</sup>

Second, CCA should be applicable also to non-actual events and actions. We are often interested in determining whether a certain merely possible event would be harmful or beneficial, were it to occur. It is not entirely clear what CCA, as formulated above, says about non-actual events. Assuming a possible worlds analysis of counterfactuals, the following two interpretations are possible:

*Interpretation 1:* A possible event  $e$  would harm (benefit) a person if and only if she is worse (better) off in the nearest  $e$ -world,  $w_e$ , than in the nearest non- $e$ -world.

*Interpretation 2:* A possible event  $e$  would harm (benefit) a person if and only if she is worse (better) off in the nearest  $e$ -world,  $w_e$ , than in the non- $e$ -world that is nearest to  $w_e$ .

If  $e$  is an actual event, the two interpretations are equivalent, since  $w_e$  is then the actual world. For non-actual events, however, the interpretations will often yield different results. Let  $e$  be the possible event of your escaping from a North Korean labour camp. According to Interpretation 1, this event would benefit you just in case you are better off in the world where you escape from the camp, than in the actual world. (Note that, for non-actual events  $e$ , the nearest non- $e$ -world is always the actual world.) According to Interpretation 2, your escape would benefit you just in case you would be better off if you were to escape, than if you were to stay in the camp. In what follows, I shall assume Interpretation 2. It is clear, I think, that this is the more reasonable interpretation. In determining whether escaping would benefit you, it is irrelevant to compare your level of well-being in the world where you escape with your level of well-being in the actual world. In the actual world, you are never (let us hope) detained in a North Korean labour camp.

Thus clarified, CCA has several virtues. It is simple, elegant, and has intuitive support. Unlike many rival accounts, it can explain why death is harmful or, in some cases, beneficial.<sup>2</sup> Suppose you die in an accident tomorrow, and that you would otherwise have had a predominantly good future life. If “on balance better off” is understood in terms of total lifetime well-being, CCA plausibly implies that the event of your death harms you.

However, CCA also faces well-known problems. Prominent among these is the “preemption problem”. Suppose that you hit me with your fist, breaking my nose. Had you not hit me

<sup>1</sup> Sometimes a distinction is made between intrinsic and extrinsic harm. (Bradley (2012), p. 392; Klockslem (2012), p. 289.) An event is intrinsically harmful to the extent that it is harmful because of its intrinsic properties. An event is extrinsically harmful to the extent that it is harmful because of its relations to other events. The same distinction can obviously be made regarding benefit. I shall understand overall harm and benefit as including both intrinsic and extrinsic harm and benefit. Thus, an event may be intrinsically harmful (beneficial) but overall beneficial (harmful), due to its effects. Similarly, an event may be extrinsically harmful (beneficial) but overall beneficial (harmful), because of its intrinsic properties.

<sup>2</sup> These and other advantages of CCA are pointed out by Jens Johansson and Olle Risberg (2017), and by Ben Bradley (2012), p. 396 f.

with your fist, I would have been hit with a baseball bat. This would have injured me even worse. In this case of what appears to be “preemptive harm”, CCA implies that you benefit me by hitting me with your fist. Many would find this conclusion plainly false. On the contrary, they would claim, you harm me. There are also parallel cases of what is seemingly “preemptive benefit”. Assume that I invite you to a party, which you enjoy very much. Had I not invited you to the party, however, you would have been invited to a concert, which you would have enjoyed even more. CCA implies that I harm you by inviting you to the party. Many would find it obvious that I do not harm you. Rather, they would maintain, I benefit you.

## 2 A Platitude and a Problem Case

The preemption problem, as usually understood, concerns the extensional adequacy of CCA. It appears that CCA misclassifies some harms as benefits, and vice versa.<sup>3</sup> I shall argue that the problems for CCA go beyond the question of extensional adequacy. As has been pointed out by others, an adequate account of harm and benefit should recognize the prudential and moral relevance of these notions.<sup>4</sup> Prudentially, harms are something we have reason to avoid, and benefits are something we have reason to pursue. Morally, harms are something we have reason not to inflict on people, and benefits are something we have reason to bestow on them.<sup>5</sup> I shall try to show that CCA fails to accord with such platitudes, in addition to not being extensionally adequate. My examples will concern prudential reasons, but they could easily be modified so as to involve also moral reasons.

The following principle is, I believe, one of the platitudes an account of harm and benefit should satisfy:

*Reason for Action:* If  $a$  and  $a'$  are alternative actions open to you in a choice situation, and doing  $a$  would benefit (not harm) you, whereas doing  $a'$  would not benefit (harm) you, then you have a prudential reason to do  $a$ , rather than  $a'$ .

It seems clear that you have reason to choose an option that would benefit you, rather than one that would harm you, or at least not benefit you. Likewise, you have reason to choose an option that would not harm you, rather than one that would.

Two remarks are called for, as regards Reason for Action. First, I am not assuming that the reason to do  $a$ , rather than  $a'$ , must be prudentially conclusive. Although I find also this assumption plausible, it is not necessary for my purposes.<sup>6</sup> Second, I rely on an intuitive

<sup>3</sup> Johansson and Risberg (2017) criticize a number of recent attempts to modify CCA, in order to solve the preemption problem. They conclude that no such modified account is likely to succeed.

<sup>4</sup> Bradley (2012), p. 394ff., includes prudential and normative importance, as well as extensional adequacy, among the desiderata for an account of harm.

<sup>5</sup> One philosopher who has been interpreted (by Bradley 2012, p. 396) as rejecting such claims is Alastair Norcross. He denies that “judgments of harm or benefit make any *distinctive* contributions to reasons for action” (Norcross 2005, p. 171, my italics). However, he goes on to state that the judgement that an action would harm a person “is certainly a relevant consideration from a consequentialist perspective...” (p. 172). Facts about harm and benefit are, according to Norcross, just facts about the relative values of the outcomes of alternative actions. (He does not discuss other events than actions.) Since the consequentialist can therefore take the relevant considerations into account without invoking the concepts of harm and benefit, these concepts play no *fundamental* (or “distinctive”) normative role, in Norcross’s view. This claim is not a denial of their normative relevance. (However, in section 5 I shall follow Bradley in arguing that a “contrastive” version of CCA, which Norcross seems to espouse, does have problems to account for the normative relevance of harm and benefit.)

<sup>6</sup> The same remark applies to the principle Reason for Preference, introduced in the next section.

understanding of what it means for something to be a reason to do one action *rather than* another. The exact analysis of such “contrastive” reason claims is controversial. In particular, it is debated whether or not contrastive reason claims can be reduced to non-contrastive ones. One reductive analysis states that a fact  $p$  is a reason to do  $a$  rather than  $a'$  iff (i)  $p$  is a reason to do  $a$  but not a reason to do  $a'$ , or (ii)  $p$  is a reason not to do  $a'$  but not a reason not to do  $a$ . According to a competing, non-reductive analysis,  $p$  is a reason to do  $a$  rather than  $a'$  iff  $p$  is a reason to do  $a$ , out of the alternative-set  $\{a, a'\}$ .<sup>7</sup> Reason for Action is eminently plausible on both analyses. If it is a fact that doing  $a$  would benefit you and doing  $a'$  would not benefit you, this fact certainly seems to be a reason to do  $a$ , but not a reason to do  $a'$ . Similarly, if it is a fact that doing  $a'$  would harm you and doing  $a$  would not harm you, this fact is a reason not to do  $a'$ , but not a reason not to do  $a$ . Further, either of these facts is, it seems, a reason to do  $a$  out of  $\{a, a'\}$ . That is, they are reasons to do  $a$ , given that  $a$  and  $a'$  are the only alternatives under consideration.

CCA is, however, incompatible with Reason for Action. Consider the following case:

*Darts:* You can throw a dart at either of two exactly similar, circular boards. Board 1 is, unlike board 2, surrounded by a much thinner circle of a different colour. You win \$1,000 if you hit board 1, \$1,200 if you hit the surrounding circle, and nothing if you miss both the board and the circle. You win \$100 if you hit board 2, and nothing if you miss it. Having more money makes you better off, and the outcomes have no other effects. You are sufficiently good at throwing darts to hit either of the boards. However, you are not good enough to hit the circle surrounding board 1 at will. If you aim for it, there is a big risk that you will miss both it and the board. In the nearest possible world where you hit board 1, it is nevertheless true that if you were not to do this, you would hit the surrounding circle. In the nearest possible world where you hit board 2, it is true that if you were not to do this, you would miss both boards, as well as the circle around board 1.<sup>8</sup>

Let  $w_{b1}$  ( $w_{b2}$ ) be the nearest world where you hit board 1 (board 2). In  $w_{b1}$  it is true that if you were not to hit board 1, you would be better off. In  $w_{b2}$  it is true that if you were not to hit board 2, you would be worse off. Hence, CCA implies that hitting board 1 would harm you, while hitting board 2 would benefit you. But you seem to have no reason to hit board 2, rather than board 1. (This holds on both of the analyses of contrastive reason claims mentioned above.) If this is correct, CCA violates Reason for Action.

If a defender of CCA does not want to reject Reason for Action, she must either deny that CCA has the implications I claim it has in Darts, or argue that you do, after all, have a reason to hit board 2, rather than board 1. The latter strategy seems doomed to fail. I cannot think of any reason whatsoever to go for board 2. Denying the purported implications of CCA may appear more promising. This could be done by questioning the plausibility of the counterfactuals assumed in Darts. I have assumed that it is true in  $w_{b1}$ , that if you were not to hit board 1, you would hit the surrounding circle. This counterfactual may be deemed implausible. It is perhaps more realistic to assume that, as seen from  $w_{b1}$ , if you were not to hit board 1, you *might* hit the

<sup>7</sup> A version of the former analysis, including only clause (i) of the *analysans*, is discussed, but not endorsed, by Justin Snedegar (2013, p. 234). Snedegar favours the latter, non-reductive analysis (2013, p. 239).

<sup>8</sup> An example with essentially the same structure as Darts was suggested, in conversation, by Jens Johansson. To transform Darts into a case involving moral reasons, we could simply assume that the well-being of other people than the agent is seriously affected by the different possible outcomes. The same goes for Coins, introduced in the next section.

surrounding circle, but might also miss both it and the board. In that case, it is not true that you *would* be better off if you were not to hit board 1. The conclusion that hitting board 1 would harm you is then avoided.

It is not at all clear, though, that the counterfactual in question is implausible. Possible worlds where you hit the surrounding circle are, in an obvious way, more similar to worlds where you hit board 1, than are worlds where you miss both the circle and the board. It seems reasonable to assume that in the nearest worlds where you miss board 1, you throw the dart in *almost* the same direction, with *almost* the same speed, etc., as you do in  $w_{b1}$ . Moreover, it suffices for the purposes of the example that there are *some* possible circumstances, however unlikely to obtain, under which the counterfactual is true. If CCA purports to be a general account of harm and benefit, it should give the right results in all metaphysically possible, or at least in all empirically possible cases.

Moreover, CCA is in trouble even if we reject the disputed counterfactual and assume, instead, that, as seen from  $w_{b1}$ , if you were not to hit board 1, you might hit the surrounding circle, and might miss both it and the board. Under this assumption, CCA implies that hitting board 1 would neither harm nor benefit you. But the implication that hitting board 2 would benefit you can hardly be avoided by a similar move. Hitting board 2 would benefit you, according to CCA, as long as it is true, in  $w_{b2}$ , that you would miss both boards, as well as the circle, were you not to hit board 2. This counterfactual is surely not implausible. The upshot is that CCA still violates Reason for Action. Since the latter principle is difficult to deny, this spells serious trouble for CCA.

### 3 Revising CCA

Perhaps a slightly modified version of CCA can avoid the problem brought out by Darts. It may be suggested that, in determining whether an intentional action,  $a$ , is harmful or beneficial, the relevant contrast world is the nearest world in which an alternative intentional action is performed, rather than a world in which the agent tries to perform  $a$ , but fails. (If so, a world where you aim at board 1 but unintentionally hit the surrounding circle cannot be the relevant contrast to the world where you succeed in hitting board 1.) This yields the following account of harm and benefit, as concerns intentional actions:

*The Nearest Alternative Account:* An action  $a$  harms (benefits) a person  $S$  if and only if  $S$  would have been better (worse) off if the agent had performed the action she would have performed, in case she had performed an alternative to  $a$ .

However, this account is no improvement of the standard CCA, as regards violations of Reason for Action. Add to Darts the assumption that there are two more boards you can hit. If you hit board 3 you win \$2000, and if you hit board 4 you win nothing. Suppose that, as seen from  $w_{b1}$ , if you were not to hit board 1, you would intentionally hit board 3. In  $w_{b2}$ , though, it is true that if you were not to hit board 2, you would intentionally hit board 4. Under these assumptions, we get the same old implausible result, viz. that hitting board 1 would harm you, while hitting board 2 would benefit you.<sup>9</sup>

<sup>9</sup> Admittedly, the conclusion that hitting board 1 would harm you is, by itself, not quite as implausible in this variant of the case as in the original version. But the conjunction of this conclusion and the conclusion that hitting board 2 would benefit you is, I submit, at least as implausible as in the original case.

At this point, a defender of CCA may suggest a somewhat more radical modification of the standard account. Instead of focusing on the action the agent would perform if he were not to act with the intention to do  $a$ , we could focus on the action that would make the recipient of harm or benefit best off, among the alternatives open to the agent in the situation.<sup>10</sup> This leads to the following account:

*The Maximizing Account:* An action  $a$  harms a person  $S$  if and only if there is an alternative action,  $a'$ , open to the agent in the situation, such that  $S$  would have been better off if the agent had done  $a'$ . An action  $a$  benefits  $S$  if and only if there is no such alternative action  $a'$ .<sup>11</sup>

In Darts, this account avoids the implausible results of the standard CCA. In the first version of Darts, hitting board 1 is the best option. (Since we assume that you cannot at will hit the surrounding circle, doing this is not among your alternatives.) The Maximizing Account thus implies, plausibly, that hitting board 1 would benefit you, and that hitting board 2 would harm you. In the revised case, hitting board 3 would make you better off than hitting either board 1 or board 2. Hence, The Maximizing Account implies, at least somewhat plausibly, that hitting board 1 and hitting board 2 would both harm you. Importantly, the account does not violate Reason for Action. It excludes the possibility that, of two alternative actions, the one that would make you better off is harmful, while the one that would make you worse off is beneficial.<sup>12</sup>

Even if we accept The Maximizing Account concerning actions, however, this is not of very much help to CCA in general. If we retain the standard account for other events, we run afoul of the following principle:

*Reason for Preference:* If possible events  $e$  and  $e'$  are alternative outcomes in a present or future situation, and  $e$  would benefit (not harm) you, were it to occur, whereas  $e'$  would not benefit (harm) you, were it to occur, then you have a prudential reason to prefer that  $e$  occurs, rather than that  $e'$  occurs.<sup>13</sup>

I take it that also Reason for Preference must be satisfied, in order for an account to reflect the prudential relevance of harm and benefit. It appears obvious that you have reason to prefer an outcome that would benefit you to one that would not. This holds *a fortiori* if the latter outcome would harm you. Also, you have reason to prefer an outcome that would not harm you to one that would.

Note, though, that Reason for Preference is restricted to pairs of alternative events. More precisely, I assume that  $e$  and  $e'$  have the same temporal location, and that the relevant  $e$ -world is identical to the relevant  $e'$ -world, up to the time at which  $e$  or  $e'$  may occur. Without this restriction, there are arguably examples where you have reason to prefer a harmful event to a beneficial one (and no reason to have the opposite preference). Let  $e$  be the event of your being rescued from a burning car at some future time  $t$ , and let  $e'$  be the event of your spraining your

<sup>10</sup> Such an account of harmful actions is proposed by Melinda Roberts (1998), p. 32; (2009), p. 203.

<sup>11</sup> This implies that  $a$  is guaranteed to benefit  $S$  if it is the only action open to the agent in the situation. To avoid this conclusion, we could add a clause to the effect that an action is beneficial only if there is some alternative that would make  $S$  worse off.

<sup>12</sup> This holds under the assumption that what actions are alternatives in a situation does not counterfactually depend on what action is performed.

<sup>13</sup> By claiming that you have a prudential reason to prefer  $e$  to  $e'$ , I just mean that there are self-interested grounds for your having this preference. I am not claiming that having the preference would make you better off than having some other preferential attitude.

ankle while jogging at  $t$ . To the extent that it makes sense to ask which of these events you now have reason to prefer,  $e'$  may appear preferable to  $e$ , even though, let us assume,  $e$  would benefit you and  $e'$  would harm you. Being rescued from a burning car entails *being* in a burning car, which is likely to involve serious physical injury and psychological trauma. In comparison, spraining an ankle is a rather trivial accident. However, a world where you are trapped in a burning car at  $t$ , and a world where you are jogging at  $t$ , cannot be identical up to  $t$ . Hence, this is not a counterexample to Reason for Preference.<sup>14</sup>

Now, consider the following case:

*Coins*: A coin-tossing mechanism selects and tosses one of two coins. If the mechanism selects coin 1, you win \$2,000 if it lands heads, and \$1,000 if it lands tails. If the mechanism selects coin 2, you lose \$2,000 if it lands heads, and \$1,000 if it lands tails. Having more money makes you better off, and the outcomes have no other effects. Let "Heads 1", "Tails 1", "Heads 2", and "Tails 2", refer to the respective outcomes. In the nearest Tails 1-world, it is true that Heads 1 would occur, were Tails 1 not to occur. Similarly, in the nearest Tails 2-world, it is true that Heads 2 would occur, were Tails 2 not to occur.

In this case, CCA implies that Tails 1 would harm you, whereas Tails 2 would benefit you. Clearly, however, you have no reason to prefer Tails 2 to Tails 1. Reason for Preference is thus violated.

Can an adherent of CCA deny the plausibility of the counterfactuals assumed in Coins? Hardly, it seems to me. Suppose that Tails 1 occurs. What would have happened otherwise? The obvious answer is Heads 1. The assumption that if coin 1 had not been tossed and landed tails, it would have been tossed but landed heads, is much more plausible than the assumption that coin 2 would or might have been tossed instead. Likewise, under the assumption that Tails 2 occurs, the natural assumption is that if coin 2 had not been tossed and landed tails, it would have been tossed but landed heads. At the very least, it should be possible to specify the details of Coins in such a way that these counterfactuals are true.

Since Coins creates serious difficulties for CCA, as an account of harmful and beneficial events that are not actions, we might ask whether The Maximizing Account can be extended to events in general. Apart from avoiding violations of Reason for Preference, such a generalized version would be a simpler and more unified account, in comparison to combining The Maximizing Account for actions with the standard CCA for other events. As applied to actions, The Maximizing Account presupposes that there is, for any action  $a$ , a set of possible actions, open to the agent in the situation, which constitute the alternatives to  $a$ . This assumption is often made in normative ethics and other areas.<sup>15</sup> In order to generalize The Maximizing Account to the domain of non-actions, we must make an analogous assumption to the effect that there is, for any event  $e$ , a set of possible events that make up the alternatives to  $e$ . Presumably, what characterizes these alternative events is that each of them could happen instead of  $e$  (in some suitable sense of 'could'). It is surely no easy task to come up with a

<sup>14</sup> Moreover, it is not clear that this example really shows that you sometimes have reason to prefer an event that would harm you, to one that would benefit you. I find it doubtful whether you have reason to prefer spraining your ankle at  $t$ , as such, to being rescued from a burning car at  $t$ , as such. Rather, you have reason to prefer to be jogging and sprain your ankle at  $t$ , as compared to being in a burning car and be rescued at  $t$ . That is, your reason for preference concerns conjunctions of a situation and an event, rather than mere events. (Of course, you may also have reason to prefer to be in the jogging situation, as such, rather than in the burning car situation, as such.)

<sup>15</sup> It is made, for example, in standard formulations of act consequentialism.

criterion for which possible events to include in such an alternative-set. Assuming that there is a solution to this problem, we might suggest the following account:

*Revised CCA:* An event  $e$  or action  $a$  harms a person if and only if there is an alternative event  $e'$  or action  $a'$ , such that she would have been better off if  $e'$  had occurred, or  $a'$  had been performed. An event  $e$  or action  $a$  benefits a person if and only if there is no such alternative event or action.

This account can handle Coins, as well as Darts. In Coins, Revised CCA implies that Heads 1 would benefit you, while the other three outcomes would all harm you. Note, however, that Revised CCA constitutes a quite substantial revision of the standard CCA. According to Revised CCA, what *would* have happened, had an event not occurred, or an action not been performed, is irrelevant to whether the event or action was harmful or beneficial. It is hence doubtful whether Revised CCA accommodates the fundamental intuition underlying the standard CCA. This intuition, I take it, is that harming means *making a difference for the worse*, as regards the subject of the harm, while benefiting means making a difference for the better.

Regardless of whether it shares the intuitive appeal of the standard account, there are reasons to be skeptical about Revised CCA. First, this account aggravates the problem of classifying some benefits as harms.<sup>16</sup> If you give a stranger \$1000 out of sheer kindness, you nevertheless harm her, according to Revised CCA, if you *could* give her \$2000. (No matter whether this is what you *would* do, if you were not to give her \$1000.) In order to avoid harming the stranger, you must do the very best for her that you can do. Similarly, if you win the second prize in a huge lottery, you are nevertheless harmed according to Revised CCA, given that you could have won the first prize.<sup>17</sup>

Second, Revised CCA appears arbitrary. According to this account, only the alternative that makes a person best off benefits her, while all other alternatives harm her. Why not claim, instead, that only the alternative that makes her worst off is harmful, while all other alternatives are beneficial? It is not the case that the former idea generally yields intuitively more plausible results than the latter. In situations where all or most alternatives are good, intuitively speaking, the latter idea corresponds better with common intuitions than the former. Thus, if giving a stranger nothing, giving her \$1000, and giving her \$2000 exhaust your alternatives in a situation, at most the first option is intuitively harmful to the stranger. Likewise, if you partake in a lottery where every ticket except one is a winning ticket, at most the alternative consisting in the non-winning ticket being drawn intuitively harms you. But if an account based on the idea that all alternatives except the best are harmful cannot be defended by appealing to the general plausibility of its implications, it is unclear how it can be defended. The idea in question can hardly be said to have strong intuitive support in itself.<sup>18</sup>

<sup>16</sup> On the other hand, it alleviates the problem of classifying some harms as benefits.

<sup>17</sup> An anonymous referee has suggested that moral reasons are tied to *wronging*, rather than to harming, and that the conclusion that you harm the stranger by giving her \$1000 is acceptable, granted that you do not thereby wrong her. I do not find this response very promising, for two reasons. First, it seems to implausibly exclude that you can wrong a person *by harming* him. If harming is, as such, morally irrelevant, wronging must always involve something over and above harming. Second, I find the claim that you harm the stranger by giving her \$1000 implausible regardless of whether harm is morally relevant. Similarly, the implausibility of the claim that winning the second prize in the lottery harms you does not presuppose the moral or prudential relevance of harm.

<sup>18</sup> Those attracted to maximizing consequentialism may perhaps find the idea inherently plausible as regards actions. Concerning events in general, however, it is difficult to see any such inherent plausibility.



In the light of these problems, it seems doubtful whether Revised CCA is an improvement of the standard account. It may be worth noting, furthermore, that restricting (rather than revising) CCA, so as to give an account of only harm or only benefit, does not help to avoid violations of Reason for Action and Reason for Preference. In Darts, CCA restricted to harm implies that hitting board 1 would harm you, while hitting board 2 would not. In Coins, we get the result that Tails 1 would harm you, whereas Tails 2 would not. Since you have no reason to hit board 2, rather than board 1, and no reason to prefer Tails 2 to Tails 1, CCA restricted to harm violates both Reason for Action and Reason for Preference. CCA restricted to benefit implies that hitting board 2 in Darts would benefit you, while hitting board 1 would not. In Coins, it follows that Tails 2 would benefit you, whereas Tails 1 would not. Again, Reason for Action and Reason for Preference are both violated.

#### 4 Contextualism to the Rescue?

In a recent paper, Justin Klockslem argues that alleged counterexamples to CCA can be rebutted by an appeal to the context-sensitivity of counterfactuals.<sup>19</sup> Although he does not discuss the problems highlighted by Darts and Coins, it may be worthwhile to consider whether his proposal can solve these problems.<sup>20</sup>

In one of the examples Klockslem uses to illustrate his proposal, you and I have separately decided to attack a third person. As it happens, I break the victim's legs before you arrive on the scene. When you discover that the victim's legs are broken, you decide to leave him alone. Had you found him uninjured, you would have killed him. Does CCA imply that my action harmed the victim, or rather that it benefited him? This depends, according to Klockslem, on the context. If we assume that I acted with bad intentions, and that I knew nothing about your plans, the relevant contrast world, to the one in which I break the victim's legs, may be one in which everybody leaves him alone. If so, my action harmed him, according to CCA. But if we instead assume that I knew about your murderous plans, and broke the victim's legs in order to save his life, the relevant contrast world is rather one in which you kill him. If so, CCA implies that my action benefited the victim.<sup>21</sup>

Klockslem finds these conclusions appealing, but one may question the plausibility of making the answer to whether an action is harmful or beneficial highly dependent on factors like the agent's intentions. Whether you are harmed by another person's action should, one might think, be determined mainly by facts involving *you*, as the subject of the putative harm. Another worry is that Klockslem invokes backtracking counterfactuals in explicating his contextualist theory. That is, he sometimes assumes that the relevant contrast world has a different non-immediate past history than the (imagined) actual world.<sup>22</sup> It is a widely held view, however, that backtracking counterfactuals are true only in quite special contexts.<sup>23</sup>

<sup>19</sup> Klockslem (2012).

<sup>20</sup> Johansson and Risberg (2017), convincingly argue that Klockslem's proposal fails to solve the preemption problem.

<sup>21</sup> Klockslem (2012), p. 295 f.

<sup>22</sup> See, e.g., the Archie and Jughead case (Klockslem 2012, p. 289).

<sup>23</sup> See Lewis (1979), p. 457; Bennett (2001), p. 187.

Furthermore, Klocksiem is not very explicit about what a context is. Judging from his examples, he seems to hold that contexts are determined partly by objective features of the case at hand, such as facts about agents' intentions, and partly by our subjective focus of attention.<sup>24</sup> In order to judge the implications of Klocksiem's view, a somewhat more precise characterization of the notion of a context would have been desirable.

Waiving these concerns, we might ask how Klocksiem's approach can be applied to Darts and Coins. In Darts, CCA satisfies Reason for Action if it implies that hitting board 1 would benefit you. This is the case if the relevant contrast world to  $w_{b1}$  is either  $w_{b2}$  or a world where you hit neither board. It could be argued that this is the most natural way to view the case. The intuition that you are benefited by hitting board 1 may be implicitly based on comparing this outcome to your hitting board 2 or neither board, rather than to the outcome where you aim at board 1, but accidentally hit the surrounding circle. In order to satisfy Reason for Preference, in Coins, it suffices for CCA to imply that Tails 1 would benefit you. This is true if the relevant contrast world, to the Tails 1 world, is one where coin 2 is tossed. Arguably, this implicit assumption underlies the intuition that you would be benefited, rather than harmed, by Tails 1.

As I have already argued, however, it cannot plausibly be denied that the circumstances in Darts and Coins could be specified in such a way that CCA has the by now familiar problematic implications. In other words, there are contexts in which CCA does imply that hitting board 1 in Darts, or the occurrence of Tails 1 in Coins, would harm you (or at least not benefit you), whereas hitting board 2, or the occurrence of Tails 2, would benefit you.

If focus of attention or salience of alternatives plays a crucial role in determining context, as Klocksiem appears to think, it is not difficult to specify such contexts. Suppose that you, as agent, and I, as spectator, know all the relevant facts in Darts, before you are offered to throw the dart. In assessing the merits of hitting board 1, we focus our attention wholly on the comparison between the consequences of successfully performing this action, and those of aiming at board 1, but accidentally hitting the surrounding circle. Likewise, in assessing the alternative of hitting board 2, we focus only on the comparison between the consequences of successfully doing so, and those of aiming at board 2, but missing it. These assumptions seem to spell out a context where the relevant contrast worlds are such that hitting board 1 would harm you, while hitting board 2 would benefit you, according to CCA.

As regards Coins, suppose that we know all the facts about the possible outcomes, before the coin-tossing mechanism is started. In determining assessing a given outcome, we attend only to the comparison between this outcome, and what would happen if the same coin were tossed, but landed with the other side up. This appears to specify a context in which CCA implies that Tails 1 would harm you, while Tails 2 would benefit you.

Hence, Reason for Action and Reason for Preference will still be violated, unless you, in the specified contexts, have a prudential reason to hit board 2 rather than board 1, or to prefer Tails 2 to Tails 1. But the context assumptions that generate the problematic implications do not seem to bring with them any such changes, regarding your prudential reasons. Under the psychological suppositions made in the last two paragraphs, you would hardly be inclined to opt for board 2 in Darts, or to prefer Tails 2 to Tails 1 in Coins. It appears, therefore, that appealing to the context-sensitivity of counterfactuals does not get CCA out of trouble.

<sup>24</sup> In the Mr. Weatherby case (Klocksiem 2012, p. 290), Klocksiem explains the supposed context-dependence partly in terms of "a shift in our attention between two different counterfactual situations ...".

## 5 Contrastivism to the Rescue?

Along with most other accounts of harm and benefit, the standard CCA construes these notions as binary relations between an event (or action) and a person. It has been suggested, however, that harm and benefit are actually three-place relations, involving an event, a person, and a contrast event.<sup>25</sup> We may be inclined to say, for example, that your hitting me with your fist, *rather than not hitting me at all*, harms me, while your hitting me with your fist, *rather than hitting me with a baseball bat*, benefits me. This may seem to indicate that there is no fact of the matter as to whether your hitting me with your fist harms me or benefits me, *simpliciter*. “Contrastivists” about harm and benefit make the general claim that events (including actions) are never harmful or beneficial *simpliciter*, for a particular person. An event is only harmful or beneficial in relation to a given contrast event.

The most straightforward contrastivist formulation of CCA seems to be the following:

*Contrastive CCA*: An event  $e$  or action  $a$ , rather than another event  $e'$  or action  $a'$ , harms (benefits) a person overall if and only if she would have been on balance better (worse) off if  $e'$  had occurred instead of  $e$ , or  $a'$  had been performed instead of  $a$ .

This account appears to avoid my main objections to CCA. As they stand, Reason for Action and Reason for Preference are not even applicable to Contrastive CCA, since they involve the notions of harm and benefit *simpliciter*. However, the following contrastive versions of these principles suggest themselves:

*Contrastivist Reason for Action*: If  $a$  and  $a'$  are alternative actions open to you in a choice situation, and doing  $a$  rather than  $a'$  would benefit you, then you have a prudential reason to do  $a$ , rather than  $a'$ .<sup>26</sup>

*Contrastivist Reason for Preference*: If possible events  $e$  and  $e'$  are alternative outcomes in a present or future situation, and  $e$  rather than  $e'$  would benefit you, then you have a prudential reason to prefer that  $e$  occurs, rather than that  $e'$  occurs.

Contrastive CCA satisfies these principles, given the following two plausible assumptions:

*Assumption 1*: If you would be on balance better-off if you were to do  $a$  than if you were to do  $a'$ , then you have a prudential reason to do  $a$ , rather than  $a'$ .

*Assumption 2*: If you would be on balance better-off if  $e$  were to occur than if  $e'$  were to occur, then you have a prudential reason to prefer that  $e$  occurs, rather than that  $e'$  occurs.

In Darts, Contrastive CCA implies that hitting board 1 rather than hitting board 2 benefits you, since the former action would make you better-off than the latter. By Assumption 1, therefore, you have a prudential reason to hit board 1 rather than hit board 2. In Coins, Contrastive CCA implies that Tails 1 rather than Tails 2 benefits you. Further, given Assumption 2, you have a prudential reason to prefer that Tails 1, rather than Tails 2, occurs.

<sup>25</sup> See Norcross (2005), p. 167 f.

<sup>26</sup> This principle was suggested by an anonymous referee.

It may thus seem as if Contrastive CCA is compatible with the prudential and moral relevance of harm and benefit. But this appearance is, I believe, largely spurious. Ben Bradley has argued that contrastive accounts of harm fail to make harm normatively important.<sup>27</sup> As Bradley points out, contrastive accounts imply that almost every performable action  $a$  has alternatives  $a'$  and  $a''$ , such that doing  $a$  rather than  $a'$  harms somebody, while doing  $a$  rather than  $a''$  benefits somebody. Since the agent cannot do  $a$  rather than  $a'$  without also doing  $a$  rather than  $a''$ , she will act harmfully as well as beneficially, whatever she does. And if harming and benefiting are both unavoidable, principles prohibiting harm will be implausible (since they cannot be obeyed), while injunctions to benefit will likely be pointless (since they cannot be disobeyed).<sup>28</sup>

Similar remarks apply, *a fortiori*, to events that are not actions. Contrastive CCA implies that, for almost any event  $e$  and person  $p$ , there are possible contrastive events  $e'$  and  $e''$ , such that  $e$  rather than  $e'$  harms  $p$ , whereas  $e$  rather than  $e''$  benefits  $p$ . Given that no event is harmful or beneficial *simpliciter*, this seems hard to reconcile with prudential or moral reasons to prevent or disprefer harmful events, and to promote or prefer beneficial ones. Although a thorough discussion of contrastive versions of CCA is beyond the scope of this paper, I therefore believe that there is reason to be skeptical about the compatibility between such accounts and the moral and prudential relevance of harm and benefit.

## 6 Conclusion

I have argued that CCA is incompatible with certain platitudinous principles, connecting harm and benefit to prudential reasons for action and preference. Since an adequate account should recognize the prudential relevance of harm and benefit, this speaks against CCA. Further, the prospects of revising or restricting CCA, so as to avoid these problems, do not seem bright.<sup>29</sup> Nor does an appeal to the context-sensitivity of counterfactuals, or to the alleged contrastive nature of harm and benefit, provide a ready solution.<sup>30</sup>

**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.

<sup>27</sup> Bradley (2012), p. 408.

<sup>28</sup> This argument is very similar to Bradley's, as I understand it.

<sup>29</sup> The arguments in this paper affect also certain accounts of harm that invoke causation as well as counterfactuals, such as the account proposed by Thomas Bontly (2016).

<sup>30</sup> Two anonymous referees for *Ethical Theory and Moral Practice* gave valuable comments on an earlier version of this paper. I wish to thank them, as well as the participants of the research seminar in practical philosophy at Uppsala University, especially Jens Johansson, Olle Risberg, Per Algander and Karl Ekendahl. Work for this article was supported by grant P14-0212:1 from the Bank of Sweden Tercentenary Foundation.

## References

- Bennett J (2001) On forward and backward counterfactual conditionals. In: Preyer G, Siebelt F (eds) *Reality and humean supervenience: essays on the philosophy of David Lewis*. Rowman & Littlefield, Lanham, pp 177–202
- Bontly TD (2016) Causes, contrasts, and the non-identity problem. *Philos Stud* 173:1233–1251
- Bradley B (2012) Doing away with harm. *Philos Phenomenol Res* 85:390–412
- Johansson J, Risberg O (2017) The preemption problem. *Philos Stud*, prepublication. <https://doi.org/10.1007/s11098-017-1019-x>
- Klocksien J (2012) A defense of the counterfactual comparative account of harm. *Am Philos Q* 49:285–300
- Lewis D (1979) Counterfactual dependence and Time's arrow. *Noûs* 13:455–476
- Norcross A (2005) Harming in context. *Philos Stud* 123:149–173
- Roberts MA (1998) *Child versus Childmaker: future persons and present duties in ethics and the law*. Rowman & Littlefield, Lanham
- Roberts MA (2009) The nonidentity problem and the two envelope problem: when is one act better for a person than another? In: Roberts MA, Wasserman DT (eds) *Harming future persons: ethics, genetics and the nonidentity problem*. Springer, Dordrecht, pp 201–228
- Snedegar J (2013) Reason claims and Contrastivism about reasons. *Philos Stud* 166:231–242