The Philosophical Quarterly Vol. 74, No. 2 *2024*
ISSN 0031-8094 https://doi.org/10.1093/pq/pqad082
Advance Access Publication 11th September 2023

PRUDENTIAL PROBLEMS FOR THE COUNTERFACTUAL COMPARATIVE ACCOUNT OF HARM AND BENEFIT

By Erik Carlson, Jens Johansson, and Olle Risberg

In this paper, we put forward two novel arguments against the counterfactual comparative account (CCA) of harm and benefit. In both arguments, the central theme is that CCA conflicts with plausible judgements about benefit and prudence.

Keywords: harm, benefit, counterfactual comparative account of harm and benefit, prudence, well-being.

The most popular account of harm and benefit is the *counterfactual comparative account (CCA)*, which can be formulated as follows:

> *CCA*: For any event *e* and person *S*, *e* harms *S* if and only if (and to the degree that) *S* would have been better off if *e* had not occurred, and *e* benefits *S* if and only if (and to the degree that) *S* would have been worse off if *e* had not occurred (e.g. Bradley 2009; Klocksiem 2012, 2022; Boonin 2014; Feit 2015, 2019, 2022, 2023; Petersson 2018; Timmerman 2019; Jedenheim Edling 2022).

Here and in what follows, 'harm' and 'benefit' should be understood in the *overall*, as opposed to *pro tanto*, sense. Correspondingly, 'better off' and 'worse off' should be understood as 'better off *on balance*' and 'worse off *on balance*', respectively.

The debate about the plausibility of CCA has primarily focused on what it says about harm, and the theory's implications concerning benefit have accordingly received less attention. In this paper, we present two arguments that seek to show that these implications are highly problematic. The central theme in both arguments is that CCA conflicts with plausible judgements about benefit and prudence. As we shall explain, these arguments are in various ways superior to related objections to CCA that have been presented in the literature.

Our first argument focuses on the following case:

> *Three Buttons*: Robin has three options: He can push button A, which will provide him with a well-being level of 100; he can push button B, which will also provide him with a well-being level of 100; and he can push button C, which will provide him with a well-being level of −100. Robin pushes B. In the nearest possible world in which he does not push B, he pushes C. In the nearest possible world in which he pushes A, it is true that if he had not pushed A, he would have pushed B. In the nearest possible world in which he pushes C, it is true that if he had not pushed C, he would have pushed B.

There are several ways in which the counterfactuals assumed in *Three Buttons* can be made plausible. For instance, it may be that A is located to the left of Robin, while B and C are located to his right, and that Robin prefers to push a button with his right hand (cf. Carlson 2020: 409).

The first thing to note about *Three Buttons* is that it is clearly permissible, prudentially speaking, for Robin to push A. For one, this verdict is highly intuitive in its own right. It also follows from this general claim, which we take to be highly appealing:

> *Claim 1*: For any action *a*, if *a* would leave the agent at least as well off as any alternative action available to her, then *a* is prudentially permissible.[1]

The second thing to note about *Three Buttons* is that according to CCA, pushing B is the only available action that would benefit Robin. In particular, pushing A would not benefit him according to CCA, since in the nearest world in which he does so, it is false that he would otherwise have been worse off. In the nearest world in which Robin pushes B, by contrast, it is true that he would otherwise have been worse off.

The implication that pushing A would not benefit Robin while another available action would do so is problematic for CCA, as it is difficult to reconcile with the intuition that pushing A is prudentially permissible. One way to see the problem is by considering the following claim:

> *Claim 2*: For any action *a*, if *a* is the only available action in the relevant choice situation that would benefit the agent, then *a* is prudentially required.

Again, CCA implies that pushing B is the only available action that would benefit Robin. Claim 2 thus implies that this action is prudentially required—and hence that all available alternatives to it are prudentially impermissible.

[1] We make the simplifying assumption that what an agent is prudentially permitted or required to do does not depend on her beliefs or on her evidence. Nothing of substance hinges on this, since in *Three Buttons* as well as *Four Buttons* (see below), we can easily add the assumption that the agent has access to all relevant information.

But this is incompatible with the claim that pushing A is prudentially permissible. In fact, if there is an action that is prudentially *required* in *Three Buttons* (in contrast to what Claim 1 yields), it seems more plausible that it is pushing A rather than pushing B. Pushing B is, in a sense, more risky than pushing A, since it brings Robin closer to the bad possible world in which he pushes C.

We take Claim 2 to be intuitively compelling on its own. It can also be supported in the following way. If *a* is the only available action in the relevant choice situation that would benefit the agent (that is, would *overall* benefit her—see above), then *a* would also be overall *better* for her than its alternatives. And it seems that if *a* would be overall better for the agent than its alternatives, then she is prudentially required to perform *a*. While an action's being *pro tanto* better for the agent than its alternatives is not sufficient for it to be prudentially required, it seems that being *overall* better for her is.

A critic of Claim 2 might suggest, however, that in some cases, it is prudentially supererogatory for an agent to benefit oneself (and to do what would be overall best for her). In particular, it might sometimes be prudentially permissible to forego a *very small* benefit. If an agent would gain one cent by effortlessly raising her arm, say, it could be denied that she is prudentially required to do so, even if there are no other beneficial options available. In *Three Buttons*, however, CCA implies that pushing B would benefit Robin *very much*. We could therefore weaken Claim 2 so as to claim that if *a* is the only option that would benefit the agent, and it would benefit her very much, then *a* is prudentially required. Moreover, our next argument only assumes that if *a* is the only beneficial option for the agent, then *a* is prudentially *permissible* (see Claim 4 below).

Our second argument against CCA focuses on the following case:

> *Four Buttons*: Robin has four options: He can push button A, which will provide him with a well-being level of 100; he can push button B, which will also provide him with a well-being level of 100; he can push button C, which will provide him with a well-being level of –100; and he can push button D, which will provide him with a well-being level of –200. Robin pushes C. In the nearest possible world in which he does not push C, he pushes D. In the nearest possible world in which he pushes A, it is true that if he had not pushed A, he would have pushed B. In the nearest possible world in which he pushes B, it is true that if he had not pushed B, he would have pushed A. In the nearest possible world in which he pushes D, it is true that if he had not pushed D, he would have pushed C.

As in *Three Buttons*, there are various ways to make the counterfactuals in *Four Buttons* plausible. For instance, the counterfactuals about pushing C and pushing D can be made plausible by assuming that A and B are located to the left

of Robin, while C and D are located to his right, and that he prefers to use his right hand rather than his left.

In *Four Buttons*, it seems evident that pushing C is prudentially impermissible. This verdict is not only intuitive in its own right but also a consequence of the following claim, which we take to be highly attractive:

> *Claim 3*: For any action *a*, if *a* would leave the agent at a negative well-being level while another available action in the relevant choice situation would leave her at a positive well-being level, then *a* is prudentially impermissible.

However, note that pushing C is the only available action that would leave Robin better off than he otherwise would have been. CCA thus implies that pushing C is the only available action that would benefit him. This result makes it hard for CCA proponents to accept the following claim:

> *Claim 4*: For any action *a*, if *a* is the only available action in the relevant choice situation that would benefit the agent, then *a* is prudentially permissible.

By CCA and Claim 4, Robin is prudentially permitted to push C. But by Claim 3, as well as intuitively, Robin is not prudentially permitted to push C. Thus, if Claims 3 and 4 are true, then CCA is false.

We take Claim 4 to be extremely plausible on its own. Moreover, the kind of support we provided for Claim 2 is even stronger when applied to Claim 4. Again, if *a* is the only available action in the relevant choice situation that would benefit the agent, then *a* would also be overall better for her than its alternatives. Surely, moreover, if *a* would be overall better for the agent than its alternatives, then *a* is least prudentially permissible. It cannot be prudentially impermissible for an agent to do what would be overall best for her.

It should be stressed that Claims 1–4 are not based on any general theory of prudence. We do not assume, for example, that these claims constitute fundamental principles, or that they specify prudentially right-making or wrong-making properties. For our arguments to be successful, it is sufficient that the claims are plausible.

It should also be stressed that our arguments only concern what CCA says about the options specified in *Three Buttons* and *Four Buttons*, respectively. We do not deny that some of CCA's other implications about these cases are plausible. For instance, given that Robin would have pushed A or B in *Four Buttons* if he had used his left hand, CCA implies that Robin's pushing a button with his right hand harms him. Even if plausible, this implication is irrelevant to the strength of our arguments.

An interesting question is whether analogues of Claims 1–4 are plausible also concerning certain attitudes.[2] Along the lines of Claims 2, for example, it might be suggested that a person is prudentially required to be glad that event *e* occurred, if *e* was the only event, of those that might have occurred in the relevant situation, which would have benefitted her. However, we are not committed to such claims about attitudes, which are much more controversial than the claims about actions that we invoke. For example, the claim just indicated could be challenged on the grounds that whether one is prudentially required to be glad that *e* occurred depends in part on what consequences that attitude would have.

The two arguments we have proposed to some extent resemble arguments against CCA that we have provided earlier (Carlson 2019, 2020; Carlson, Johansson, and Risberg 2021, 2023a).[3] We believe, however, that our new arguments are more compelling for several reasons. First, our earlier main normative objection to CCA was that it violates a certain principle, which we called '*Reason for Action*', concerning reasons for performing one action rather than another (Carlson 2019: 797, 2020: 408; Carlson, Johansson, and Risberg 2021: 169). Roughly, *Reason for Action* says that if option *a* would benefit the agent and option *b* would harm her, then she has a reason to perform *a* rather than *b*. It could be argued, however, that this is false if there is a third option, *c*, which would be even more beneficial for the agent than *a*. In such a case, one might claim, the agent only has reasons to perform *c*, and hence no reason to perform *a* rather than *b*. Moreover, an essential part of our earlier arguments, when translated into the terms of *Four Buttons*, is the claim that Robin has *no* reason, however weak, to push button C rather than button B. It could be argued, however, that one such reason is that in the nearest world in which Robin pushes C, unlike in the nearest world in which he pushes B, it is true that he would have been worse off if he had acted otherwise. Furthermore, 'negative existential intuitions' about reasons of the kind on which our earlier arguments rely—that is, intuitions with contents of the form *there is no reason at all for* S *to do* a—have sometimes been taken to be unreliable (see Schroeder 2007: 92–8). Since our new arguments make no assumptions about reasons (let alone reasons to perform one action *rather than* another), they avoid these and similar objections.

Second, our new arguments also avoid a potential problem with the main case, called '*Darts*', that Carlson (2019: 798; 2020: 408) employs in order to show that CCA violates *Reason for Action*. In this case, the agent's actual action

[2] This issue was raised by an anonymous referee.

[3] *Three Buttons* and *Four Buttons* draw on the '*Buttons*' case, appearing in Carlson (2020: 409) and Carlson, Johansson, & Risberg (2023a). Feit (2022, 2023) argues that our earlier arguments can be met by an appeal to 'plural harm'. For a rejoinder, see Carlson, Johansson, & Risberg (2023a). This rejoinder would also apply to a criticism of our new arguments analogous to Feit's criticism of our earlier arguments.

is assumed to be such that, if she had not performed it, she would not have performed any of the alternative actions available to her but would instead have done something that is not even an option for her. It is questionable whether such situations are possible.[4] Clearly, no such assumption is made in *Three Buttons* or *Four Buttons*.

Third, our earlier arguments against CCA rely on assumptions about harm as well as about benefit. By contrast, as already indicated, our new arguments invoke no assumptions about harm. In this respect, our arguments are simpler and proceed from a sparser base of intuitions.

It may be suggested, however, that our focus on benefit is poorly motivated since analogous arguments can be stated in terms of harm.[5] Consider a case that is exactly like *Three Buttons*, except that Robin's pushing A would provide him with a well-being level of −100, pushing B would also provide him with a well-being level of −100, and pushing C would provide him with a well-being level of 100. According to CCA, pushing B is the only harmful option in this choice situation. Furthermore, according to a mirror image of Claim 2, if there is only one harmful option in a choice situation, then only this option is prudentially impermissible. But it seems clear that it is also prudentially impermissible for Robin to push A. However, while this is indeed an argument in terms of harm that seems analogous to ours, it is not a good argument against CCA. The problem is that the relevant mirror image of Claim 2 is not intuitively plausible. If there is a very beneficial option available, another option may, it seems, be prudentially impermissible even if it is not harmful.

Of course, there is another, much more plausible claim about harm that might also be regarded as a mirror image of Claim 2—namely that if there is only one harmful option in a choice situation, then this option (although not necessarily only this option) is prudentially impermissible. However, this latter claim creates no problem for CCA in the present variant of *Three Buttons*. In particular, in this case, it is intuitively correct that pushing B—again, the only harmful option according to CCA—is prudentially impermissible.

It should be clear that similar remarks apply to *Four Buttons* as well. It does not seem, then, that our arguments can be recast in terms of harm.

In light of our focus on benefit, one reaction to our arguments might be to abandon what CCA says about benefit but preserve what it says about harm. However, this is incompatible with the intuitively plausible and widely accepted view that harm and benefit should be given a parallel treatment. Moreover, assuming that an event cannot both harm and benefit a person overall, CCA about harm implies that an event benefits a person only if she

[4] For a related criticism of Carlson's argument, see Klocksiem (2019).
[5] This suggestion was made by an anonymous referee.

would not have been better off had the event not occurred. The correct account of benefit must therefore resemble CCA. Granted, the implication allows that a person is benefitted by an event that leaves her equally well off, or neither harmed nor benefitted by an event that leaves her better off than she would otherwise have been. But a proponent of CCA about harm who endorses one or both of these possibilities must explain why a person cannot be *harmed* by an event that leaves her equally well off, or neither harmed nor benefitted by an event that leaves her *worse* off than she would otherwise have been. These asymmetries between harm and benefit appear implausible (see also Carlson, Johansson, and Risberg 2023b: 559–60).

Another reaction to our arguments would be to claim that CCA's incompatibility with Claims 1–4 shows not that CCA should be rejected but that Claims 1–4 should be rejected. Of course, the plausibility of this suggestion depends, among other things, on what can be said for and against CCA in other respects—an issue that is much too wide to be discussed here. However, given the attractiveness of Claims 1–4, even ardent proponents of CCA should accept that its incompatibility with those principles constitutes, at the very least, a challenge that needs to be addressed.[6]

REFERENCES

Boonin, D. (2014) *The Non-Identity Problem and the Ethics of Future People*. New York, NY: OUP.
Bradley, B. (2009) *Well-Being and Death*. New York, NY: OUP.
Carlson, E. (2019) 'More Problems for the Counterfactual Comparative Account of Harm and Benefit', *Ethical Theory and Moral Practice*, 22/4: 795–807.
———. (2020) 'Reply to Klocksiem on the Counterfactual Comparative Account of Harm', *Ethical Theory and Moral Practice*, 23/2: 407–13.
Carlson, E., Johansson, J., and Risberg, O. (2021) 'Well-Being Counterfactualist Accounts of Harm and Benefit', *Australasian Journal of Philosophy*, 99/1: 164–74.
——— (2023a) 'Benefits Are Better than Harms: A Reply to Feit', *Australasian Journal of Philosophy*, doi: 10.1080/00048402.2022.2154814.
———. (2023b) 'Plural Harm: Plural Problems', *Philosophical Studies*, 180/2: 553–65.
Feit, N. (2015) 'Plural Harm', *Philosophy and Phenomenological Research*, 90/2: 361–88.
———. (2019) 'Harming by Failing to Benefit', *Ethical Theory and Moral Practice*, 22/4: 809–23.
———. (2022) 'How Harms Can Be Better than Benefits: Reply to Carlson, Johansson, and Risberg', *Australasian Journal of Philosophy*, 100/3: 628–33.
———. (2023) *Bad Things: The Nature and Normative Role of Harm*. New York, NY: OUP.
Jedenheim Edling, M. (2022) 'A New Principle of Plural Harm', *Philosophical Studies*, 179/6: 1853–72.
Klocksiem, J. (2012) 'A Defense of the Counterfactual Comparative Account of Harm', *American Philosophical Quarterly*, 49/4: 285–300.
Klocksiem, J. (2019) 'The Counterfactual Comparative Account of Harm and Reasons for Action and Preference: Reply to Carlson', *Ethical Theory and Moral Practice*, 22/3: 673–7.

[6] We are grateful to Magnus Jedenheim Edling and several anonymous referees for their helpful comments. Erik Carlson and Jens Johansson's work on this paper was supported by Grant 2018-01361 from Vetenskapsrådet and Grant P21-0462 from Riksbankens Jubileumsfond. Olle Risberg's work on this paper was supported by Grant 2020-01955 from Vetenskapsrådet.

———. (2022) 'Harm, Failing to Benefit, and the Counterfactual Comparative Account', *Utilitas*, 34: 428–44.
Petersson, B. (2018) 'Over-Determined Harms and Harmless Pluralities', *Ethical Theory and Moral Practice*, 21/4: 841–50.
Schroeder, M. (2007) *Slaves of the Passions*. Oxford: OUP.
Timmerman, T. (2019) 'A Dilemma for Epicureanism', *Philosophical Studies*, 176/1: 241–57.

Uppsala University, Sweden