CAUSAL ACCOUNTS OF HARMING

BY

ERIK CARLSON, JENS JOHANSSON, AND OLLE RISBERG

Abstract: A popular view of harming is the causal account (CA), on which *harming is causing harm*. CA has several attractive features. In particular, it appears well equipped to deal with the most important problems for its main competitor, the counterfactual comparative account (CCA). However, we argue that, despite its advantages, CA is ultimately an unacceptable theory of harming. Indeed, while CA avoids several counterexamples to CCA, it is vulnerable to close variants of some of the problems that beset CCA.

1. Introduction

A popular idea in the debate on the nature of harming is that *harming is causing harm*. More precisely,

The causal account of harming (CA)
An event *e* harms a person *S* if and only if there is a state of affairs *d* such that (i) *e* causes *d* to obtain, and (ii) *d* is a harm for *S* (e.g. Shiffrin 1999, 2012; Suits 2001; Harman 2004, 2009; Velleman 2008; Hanser 2009, 2019; Thomson 2011; Smuts 2012; Gardner 2015, 2016, 2017, 2019a, 2019b; Northcott 2015; Rabenberg 2015; Bontly 2016).

Different advocates of CA specify clause (ii) in different ways. On one view, what it takes for a state of affairs *d* to be a harm for *S* is that it is intrinsically bad for *S*; on another view, what it takes is instead that *S* would have been better off if *d* had not obtained. Yet other views about (ii) exist, yielding a great number of versions of CA.

Even without looking at the details, however, it seems clear that CA has several attractive features. To begin with, many clear cases of harming intuitively involve causation. Moreover, CA appears to be well equipped

Pacific Philosophical Quarterly •• (2021) ••–•• DOI: 10.1111/papq.12390

to deal with the most important problems for its main competitor in the literature on harming:

The counterfactual comparative account of harming (CCA)
An event *e* harms a person *S* if and only if *S* would have been better off if *e* had not occurred (e.g. Parfit 1984: 69; Feit 2002, 2015, 2016, 2019; Bradley 2004, 2009; Klocksiem 2012; Boonin 2014; Timmerman 2019).

One problem with CCA concerns *preemption* (Norcross 2005; Bradley 2012; Gardner 2015; Hanna 2016; Johansson and Risberg 2019). Here is a typical case:

Tear Gas. The Joker sprays tear gas in exactly one of Batman's eyes. If he had not done that, he would have sprayed tear gas in both of Batman's eyes, which would have made Batman even worse off. One of the alternatives available to the Joker, however, was to simply leave Batman alone.

Intuitively, the Joker's action harms Batman, but CCA entails that it does not. CA, by contrast, seems capable of handling this case. Plausibly, the Joker's action causes some state of affairs that satisfies (ii) to obtain, such as *Batman's being in pain*.

Another problem concerns *creation*. Consider this case:

Misery. Bamm-Bamm is created. Unfortunately, his life is altogether miserable. If he had not been created, he would have never existed at all.

It seems clear that being created harms Bamm-Bamm. Hence, if CCA is true, Bamm-Bamm would have been better off if he had not been created. But that in turn implies that Bamm-Bamm would have occupied a well-being level even if he had never existed. That assumption is far from unproblematic (e.g. Carlson and Johansson 2018). Proponents of CA, on the other hand, need no such assumption in order to accommodate the judgment that being created harms Bamm-Bamm. For instance, they can simply say that *Bamm-Bamm's having a miserable life* is a harm for him, and that creating him causes it to obtain.

The preemption and creation problems indicate that CCA undergenerates harming. The *failure to benefit* problem indicates that CCA also overgenerates harming. Consider Ben Bradley's much discussed case (Bradley 2012: 397; Feit 2019; Purves 2019; Johansson and Risberg 2020, forthcoming):

Golf Clubs. Batman contemplates giving a set of golf clubs to Robin, but eventually decides to keep them. If he had not decided to keep them, he would have given the clubs to Robin, which would have made Robin better off.

CCA yields that Batman's decision harms Robin. Intuitively, however, the decision merely fails to benefit Robin; it does not harm him. Once again, CA seems less vulnerable to the problem. Advocates of CA can deny that the case even involves any state of affairs that is a harm for Robin; for example, *Robin's lacking golf clubs* is not intrinsically bad for Robin. And even assuming that *Robin's lacking golf clubs* is a harm for Robin, Batman's decision arguably does not *cause*, but merely *allows*, this state of affairs to obtain (Purves 2019).

A more rarely noted reason that CCA apparently overgenerates harming concerns what we might call *mere indicators*: harmless events that are reliable indicators of harming. Here is one example:

Ouch. Wilma feels intense pain, and says 'ouch' as a result. If she had not said 'ouch', that would have been because she would not have felt any pain.

Because Wilma would have been better off if she had not said 'ouch', CCA implies that her saying 'ouch' harms her. That seems wrong. This problem, too, is no embarrassment to CA. While *Wilma's feeling intense pain* is a harm for her, her saying 'ouch' does not cause it to obtain.[1]

Because of these advantages, an evaluation of the various possible versions of CA is called for. The purpose of this paper is to provide such an evaluation. We shall argue that, despite its advantages, CA is ultimately an unacceptable theory of harming. At least, on the most popular and natural ways of specifying clause (ii), CA faces problems that are no less serious than those that beset CCA. Indeed, while CA as we have seen avoids several counterexamples to CCA, it will be shown to face close variants of some of these problems.

After some preliminary remarks (Section 2), we consider three general approaches to how clause (ii) in CA should be understood. The first is to understand it in terms of *intrinsic badness* (Section 3) and the second in *temporal* terms (Section 4). While these two approaches have already been criticized in the literature, our criticism is, for reasons that will become clear, more difficult to resist. Our criticism of these two approaches also reflects several ways in which the third approach, which is to understand (ii) in *counterfactual* terms, is the most promising one. Like the other two, however, this approach can be shown to face counterexamples of various kinds (Sections 5–8). We end with some concluding remarks, where we among other things present a diagnosis of why CA fails and note some further issues for which our results are important (Section 9).

2. *Preliminaries*

Some stage setting is in order. First, CA and its competitors are typically intended not as mere extensional claims, but as claims about *what it is* for

an event to harm someone. Thus, while we formulate CA and other accounts of harming as instances of the claim schema '*p* iff *q*', we mean these instances to be shorthand for instances of 'what it is for it to be the case that *p* is for it to be the case that *q*'. Nothing in our discussion depends on this, however, because our arguments, if successful, show that CA is not even extensionally adequate.

Second, CA requires that events can cause states of affairs to obtain (unless no event harms anyone). We shall simply assume that this is the case. Moreover, as is standard, we shall assume that states of affairs can be logically complex, by involving negation, conjunction, and so on. Other than that, we shall make no particular assumptions about how states of affairs should be understood.

Third, we shall not presuppose any specific theory of causation, but intend our various causal judgments to be ones that any plausible theory should accommodate. That said, it is worth noting that if CA is to have the advantages over CCA suggested above, one likely has to assume that simple counterfactual theories of causation are false. In *Tear Gas*, for example, one cannot consistently say both that the Joker's spraying tear gas in exactly one of Batman's eyes causes *Batman's being in pain* to obtain and that *e* causes *d* to obtain only if *d* would not have obtained if *e* had not occurred.

Fourth, we shall make the plausible assumption that parallel accounts should be given of harming and benefiting. This means, to begin with, that an adherent of CA is also committed to the following causal account of benefiting:

The causal account of benefiting
An event *e* benefits a person *S* if and only if there is a state of affairs *d* such that (i) *e* causes *d* to obtain and (ii) *d* is a benefit for *S*.

Furthermore, clause (ii) in the causal account of benefiting should be given an analysis parallel to (ii) in CA. For instance, if *d* is a harm for *S* just in case *d* is intrinsically bad for *S*, then *d* is a benefit for *S* just in case *d* is intrinsically good for *S*. Both adherents and opponents of CA typically accept that harming and benefiting are in this way analogous (see e.g. Gardner 2016, p. 330).

Fifth, while there are different views about what it takes for a state of affairs to be a harm (i.e. to satisfy CA's condition [ii]), we shall focus on views that assume an essential connection between harm and well-being. However, nothing in what follows hinges on this – our arguments can easily be reformulated so as to apply also to versions of CA that assume other views about harm.

Sixth, a distinction is commonly made between *pro tanto* harming and overall harming. Suppose that a doctor gives her patient a medicine that cures him from a terrible disease but has the side effect of making him feel

pain in his foot. The doctor's action, we may suppose, benefits the patient overall. Even so, one may want to insist that it is harmful *in some respect*, that is, *pro tanto* harmful, due to the pain it causes the patient to feel. Given this reasonable distinction, CA is as it stands best understood as an account of *pro tanto* harming. (Otherwise, it would incoherently imply that the doctor's action is both overall harmful and overall beneficial to the patient.) However, CA can easily be expanded into a theory also of overall harming. In particular, it is plausible to assume that *e* harms *S* overall if and only if the sum of the degrees to which *e pro tanto* harms *S* is greater than the sum of the degrees to which *e pro tanto* benefits *S* (and analogously for overall benefiting).[2] The degrees to which *e pro tanto* harms *S* is then, at least for adherents of CA, naturally taken to depend on the degrees of the harms it causes for *S* (and analogously for degrees of *pro tanto* benefiting).[3]

How claims about degrees of harm should be understood in turn depends on how (ii) is understood. However, for each understanding of (ii) that we shall discuss, it will be fairly straightforward (and therefore left implicit) how degrees of harm are most plausibly determined. For instance, if *d* is a harm for *S* just in case *d* is intrinsically bad for *S*, then surely the degree to which *d* is a harm for *S* equals the degree to which *d* is intrinsically bad for *S*.

3. *The causal-intrinsic badness account*

The simplest version of CA is likely the one to which we just alluded. It can be formulated as follows (cf. Suits 2002; Smuts 2012):

The causal-intrinsic badness account of harming (CIBA)
Harming: An event *e* harms a person *S* if and only if there is a state of affairs *d* such that (i) *e* causes *d* to obtain, and (ii) *d* is a harm for *S*.
Harm: A state of affairs *d* is a harm for *S* if and only if *d* is intrinsically bad for *S*.

CIBA yields intuitively plausible results in a range of cases, including *Tear Gas* (Section 1). Given that the Joker's action causes Batman pain, and that pain is intrinsically bad for Batman, CIBA straightforwardly entails that the action harms Batman.

The most obvious problem with CIBA concerns death. Unless we make the highly controversial assumption that being dead can be *intrinsically* bad for the deceased, CIBA yields that the event of death cannot harm the one who dies. Because this is implausible, CIBA is itself implausible. Unfortunately, this argument will not move those who deny that death ever harms its victim; indeed, some of them base their denial on something like CIBA (e.g. Suits 2002; Smuts 2012). It would not be dialectically ideal, then, for our criticism of CIBA to rely on the harmfulness of death.

However, CIBA fails for other reasons. Consider this case:

Happiness Reduction. Batman is extremely happy. The Joker gives him a pill that causes his happiness to diminish significantly, making him merely moderately happy.

Intuitively, the Joker's action harms Batman, both *pro tanto* and overall, even though it does not cause him anything intrinsically bad. However, CIBA implies that it does not even *pro tanto*, let alone overall, harm Batman. While this sort of problem for views like CIBA has been noted in the literature (e.g. Gardner 2015, p. 431; Rabenberg 2015, pp. 5–6), there is a nearby and perhaps even more serious problem that, as far as we know, has not been pointed out. Let us add to the case that the Joker's action also causes some minor intrinsic good for Batman – maybe, for example, the taste of the pill is mildly pleasant. (Note that unless a dead person can have experiences, the corresponding move cannot be made with regard to the death case.) Assuming the outlined connection between *pro tanto* and overall harming and that harming and benefiting should be given parallel accounts (Section 2), CIBA yields that the Joker's action *benefits* Batman *overall*. This is an unacceptable result.

The CIBA thus undergenerates *pro tanto* harming and overgenerates overall benefiting. The following case shows that it also undergenerates *pro tanto* benefiting and overgenerates overall harming:

Painkiller. Batman feels intense pain in his upper body, but Robin gives him a painkilling substance by inserting a syringe in his leg. While the sting from the syringe causes Batman some slight discomfort, the painkilling substance relieves the pain in his upper body entirely.

In his discussion of a relevantly similar case, Michael Rabenberg (2015, p. 7) suggests that the action does not harm the person at all. However, due to the discomfort caused by the sting, it seems plausible that Robin's action does harm Batman *pro tanto*, just as CIBA implies. The problems for CIBA lie elsewhere. To begin with, because Robin's action does not cause anything intrinsically good for Batman, an account of benefit parallel to CIBA implies that Robin's action does not benefit Batman, even *pro tanto*. Moreover, CIBA also implies that Robin's action harms Batman overall (assuming, again, the connection between *pro tanto* and overall harming outlined in Section 2). Both these results are unacceptable.

It may seem that CIBA's account of harm can be modified so as to handle these objections. Consider the following:

Revised CIBA
Harming: An event *e* harms a person *S* if and only if there is a state of affairs *d* such that (i) *e* causes *d* to obtain, and (ii) *d* is a harm for *S*.
Harm: A state of affairs *d* is a harm for *S* if and only if *d* is intrinsically bad for *S*, or the negation of *d* would be intrinsically good for *S*.

The parallel account of benefit, then, says that *d* is a benefit for *S* just in case *d* is intrinsically good for *S*, or the negation of *d* would be intrinsically bad for *S*.

The explanation why Batman is benefited in *Painkiller*, it may be argued, is that Robin's action causes a state of affairs whose negation would be intrinsically bad for Batman, such as *Batman's not feeling intense pain in his upper body after time* t, to obtain. Hence, Revised CIBA correctly implies that Robin's action *pro tanto* benefits Batman. Moreover, the account implies that Batman is also benefited overall, because *Batman's feeling slight discomfort from the sting of the syringe* is intrinsically better for him than *Batman's feeling intense pain in his upper body after time* t.

In *Happiness Reduction*, one might claim that the Joker's action causes states of affairs whose negations would have been intrinsically good for Batman, such as *Batman's not being extremely happy after* t, to obtain. If so, Revised CIBA can deliver the intuitively correct result that the Joker's action harms Batman overall. Similarly, if the Joker kills Batman, he causes states of affairs like *Batman's not being happy after* t to obtain.[4] It thus seems that Revised CIBA can also account for the harmfulness of death.

Conee (2006, pp. 183–185), Bradley (2012, p. 409), and Feit (2015, pp. 365–366) briefly discuss an account of harming that resembles Revised CIBA, although one on which harming is understood as causing something intrinsically bad to obtain or causing something intrinsically good *not* to obtain, rather than as causing something intrinsically bad to obtain or causing the negation of something intrinsically good to obtain. While Bradley and Feit reject this account, Conee tentatively accepts it, at least with respect to the harmfulness of death. Given the plausible assumption that causing a state of affairs not to obtain is equivalent to causing its negation to obtain, the difference between Revised CIBA and the account discussed by Conee, Bradley, and Feit is merely verbal. Bradley and Feit both stress the difficulty of determining what states of affairs an event causes not to obtain. Moreover, Bradley is skeptical about whether the account avoids the preemption problem. He refers to a case in which Bobby Knight 'chokes a philosopher, injuring her windpipe; if he hadn't choked her, he would have torn her arms off, which would have been much worse for her' (2012, p. 407).[5] Bradley claims that Knight's choking the philosopher 'causes him not to rip off her arms' (2012, p. 409), and hence that the Revised-CIBA-like account implausibly entails that Knight's action overall benefits the philosopher.

Bradley does not motivate why a defender of the account must agree that Knight's choking the philosopher causes him not to rip off her arms. Making this causal claim without argument is in tension with his assertion that it is hard to say what states of affairs an event causes not to obtain. We believe, however, that a more convincing argument, also involving preemption, can be made against Revised CIBA. Consider the following case:

Tear Gas Again. Riddler is about to spray tear gas in Batman's left eye. The Joker can prevent this, either by simply telling Riddler to leave Batman alone, or by spraying tear gas in Batman's right eye. (Riddler and the Joker have agreed to leave at least one of Batman's eyes undamaged.) If the Joker tells Riddler to leave Batman alone, no tear gas will be sprayed. If Batman gets tear gas in his left eye he will suffer 15 units of pain. If he gets tear gas in his right eye he will suffer 10 units of pain.

Intuitively, the Joker's telling Riddler to leave Batman alone would benefit Batman overall, while spraying would harm him overall. If Revised CIBA is to imply that telling Riddler to leave Batman alone would be beneficial, we must assume that it would cause *Batman's not suffering 15 units of pain* to obtain. This seems plausible. But if the Joker's telling Riddler to leave Batman alone would cause this state of affairs to obtain, so would his spraying. No matter what theory of causation we accept, it would be very implausible to claim that the Joker's two alternatives differ with respect to causing *Batman's not suffering 15 units of pain* to obtain. Either both actions would cause this state to obtain, or neither action would. If the Joker's spraying would cause *Batman's not suffering 15 units of pain* to obtain, Revised CIBA implies that this action would benefit Batman overall. Although it would cause *Batman's suffering 10 units of pain* to obtain, and hence be *pro tanto* harmful, this harm is outweighed by the benefit constituted by *Batman's not suffering 15 units of pain*. Hence, Revised CIBA seems incompatible with the conjunction of the two claims, that the Joker's telling Riddler to leave Batman alone would overall benefit Batman, and that the Joker's spraying would overall harm Batman.

4. *Causal-temporal accounts*

Another possible way to overcome at least some of CIBA's problems is to include a temporal condition in the account of harm, to the effect that a harm always involves a decrease in a person's level of well-being. In *Happiness Reduction*, for example, it seems plausible that the reason why the Joker's action harms Batman is that it makes Batman be less well off after the action than he was before. Let us call an account of harming *temporal* just in case it essentially involves a comparison between a person's well-being levels before and after the occurrence of the relevant event.

A temporal account need not, but arguably should, involve causation. It would be hopelessly implausible to claim that an event harms a person just in case she is worse off after the occurrence of the event than she was before it. This view vastly overgenerates harming, as it entails that all events that occur at the same time as a harmful event are also harmful. A better non-causal temporal account is counterfactual:

The counterfactual-temporal account (COTA)
An event *e*, occurring at a time *t*, harms a person *S* if and only if *S* is worse off after *t* than *S* was before *t*, and this would not have been the case if *e* had not occurred.

COTA is, however, just as vulnerable as CCA to the preemption problem. Consider again *Tear Gas* (Section 1). Assuming that the tear gas in Batman's eye makes him worse off after the Joker's action than he was before, and that having tear gas in both eyes would have made him worse off still, COTA counterintuitively entails that the Joker's spraying tear gas in exactly one of Batman's eyes does not harm him.

This problem is avoided if we instead choose a corresponding causal condition (e.g. Perry 2003; Velleman 2008; Foddy 2014):

The first causal-temporal account (CATA-1)
Harming: An event *e*, occurring at a time *t*, harms a person *S* if and only if there is a state of affairs *d* such that (i) *e* causes *d* to obtain immediately after *t*, and (ii) *d* is a harm for *S*.
Harm: A state of affairs *d* is a harm for *S* if and only if it consists in *S*'s being worse off than *S* was earlier.

The fact that Batman would have been even worse off after *t*, if the Joker had sprayed tear gas in both of his eyes, does not exclude that what the Joker actually does causes Batman's well-being level to decline after *t*.

Like COTA above (and CATA-2 below), CATA-1 implies that an event harms someone only if she undergoes a decrease in well-being. This implication is disputable for several reasons. One is that it entails that in *Misery* (Section 1), Bamm-Bamm is harmed only if he occupied a well-being level before he came into existence. The assumption that a person occupies a well-being level before she exists is at least controversial and, we think, not particularly natural, but we shall not pursue this complicated issue further here. Cases like the following are also known to cast doubt on the claim that an event harms someone only if she undergoes a decrease in well-being (e.g. Rabenberg 2015, p. 18):

Hampered Recovery. At *t*, Bamm-Bamm is recovering from a long period of illness. If nobody interferes, he will recover fully, and his well-being level after *t* will be much higher than it was before *t*. Unfortunately, Bamm-Bamm's mother Betty suffers from Münchausen by proxy, and does not want Bamm-Bamm to fully recover. She therefore gives him a drug at *t*, which causes some of his symptoms to become chronic. Bamm-Bamm's well-being level is never lower than it was at any earlier time, but as a result of Betty's action, his well-being level after *t* is the same as it was before *t*.

It seems that Betty's action harms Bamm-Bamm, even though he undergoes no decrease in well-being.[6] Some writers are unconvinced, however. For example, Bennett Foddy (2014, p. 160) suggests that it is acceptable to say that an action like Betty's is harmless, because its being a prevention

of benefit suffices to explain why it is morally objectionable. Stephen R. Perry (2003, p. 1298), David Velleman (2008, p. 243), and Judith Jarvis Thomson (2011, pp. 444–445) suggest, instead, that this sort of case actually does involve a decrease in well-being, because the victim's *chances* of recovering decrease. While we believe that these responses can be shown to be unsuccessful (cf. Rabenberg 2015, p. 18), there is no need to enter into this controversy here. For CATA-1 fails for a simple – although as far as we know, not yet noted – reason that is independent of the *Hampered Recovery* problem.

Consider the following case:

Pain Increase. The Joker sprays tear gas in Batman's eyes at time *t*. Simultaneously, Riddler sprays a pain increasing chemical in Batman's eyes. Unlike the tear gas, the chemical would not have been painful on its own. (Batman thus would not have felt any pain if the Joker had not sprayed the tear gas.) However, the chemical makes the pain caused by the tear gas even worse.

Riddler's action plausibly does not cause Batman to be worse off after *t* than he was before *t*. (The latter is something that the Joker's action causes all by itself, and is not overdetermined.) But Riddler's action still harms Batman, given that he would have been less badly off after *t* if Riddler had acted differently. So CATA-1 is false.

The following account handles this case better:

The second causal-temporal account (CATA-2)
Harming: An event *e*, occurring at a time *t*, harms a person *S* if and only if there is a state of affairs *d* such that (i) *e* causes *d* to obtain immediately after *t*, and (ii) *d* is a harm for *S*.
Harm: A state of affairs *d* is a harm for *S* if and only if it consists in *S*'s occupying a particular well-being level *l*, such that *l* is lower than the well-being level *S* occupied immediately before *S* occupied *l*.

Plausibly, Riddler's action in *Pain Increase* causes Batman to occupy a particular well-being level after *t*. If so, CATA-2 entails that it harms Batman.

But CATA-2 fails too. Consider

Pain Relief. Fred suffers from a painful condition that will inevitably make his well-being level after *t* lower than it was before *t*. At *t* Wilma gives Fred a drug that will to some extent relieve the future pain caused by Fred's condition.

It seems clear that Wilma's action benefits Fred and does not harm him, even *pro tanto*. However, on the plausible assumption that the action causes Fred to occupy a particular well-being level after *t*, CATA-2 (unlike CATA-1) entails that it does harm him.

5. *Causal-counterfactual accounts*

Recently, a number of writers have proposed what might be called *causal-counterfactual* accounts of harming (e.g. Thomson 2011; Gardner 2015, 2016, 2017, 2019a, 2019b; Northcott 2015; Bontly 2016). For reasons that will emerge, this is likely the most promising kind of causal theory of harming. The simplest causal-counterfactual account can be formulated as follows:

The simple causal-counterfactual account (SCA)
Harming: An event *e* harms a person *S* if and only if there is a state of affairs *d* such that (i) *e* causes *d* to obtain, and (ii) *d* is a harm for *S*.
Harm: A state of affairs *d* is a harm for *S* if and only if *S* would have been better off if *d* had not obtained.

SCA seems to yield plausible verdicts in the cases discussed so far. To begin with, SCA (like most other versions of CA) avoids the counterexamples to CCA noted in Section 1. In *Tear Gas*, the state of affairs of *Batman's being in pain*, for example, is caused to obtain by the Joker's action, and Batman would have been better off if it had not obtained. Unlike CCA, then, SCA implies that the Joker's action harms Batman. In *Misery*, there are plausibly several states of affairs *d* such that if *d* had not obtained, then Bamm-Bamm would have been better off. Because Bamm-Bamm's being created surely causes at least some of those states of affairs to obtain, SCA yields that his being created harms him, which is the intuitively correct result. In *Golf Clubs*, there are admittedly plenty of states of affairs involved that satisfy SCA's *Harm* component; for instance, Robin would have been better off if *Robin's lacking golf clubs* had not obtained. However, SCA's proponents can justifiably deny that Batman's decision *causes*, rather than merely *allows*, any such state of affairs to obtain (Purves 2019). As for *Ouch*, Wilma's saying 'ouch' does not cause *Wilma's being in pain* – or, it seems, any other state of affairs satisfying SCA's *Harm* component – to obtain. In both *Golf Clubs* and *Ouch*, then, SCA, unlike CCA, avoids the implication that the relevant event harms the person.

Furthermore, SCA also seems to handle the counterexamples to the accounts discussed in the previous two sections. Recall, for example, this counterexample to CIBA (Section 3):

Happiness Reduction. Batman is extremely happy. The Joker gives him a drug that causes his happiness to diminish significantly, making him merely moderately happy.

Let *d* be *Batman's undergoing a diminishing of his happiness*. The Joker's action causes *d* to obtain, and Batman would have been better off if *d* had

not obtained – he would have remained extremely happy. Unlike CIBA, then, SCA yields the intuitive verdict that the Joker's action harms Batman.

Recall also the counterexample to CATA-1, *Pain Increase*, where Riddler sprays a chemical in Batman's eyes that intensifies the pain caused by the Joker's action (Section 4). Riddler's action clearly harms Batman, and that is also what SCA implies. For instance, Batman would have been better off if *Batman's having the chemical in his eyes* had not obtained; and Riddler's action causes this state of affairs to obtain.

Before going into problems with SCA, let us note that an influential, slightly more complex causal-counterfactual account has been proposed by Molly Gardner (2015, p. 434):

Gardner's causal-counterfactual account (GCA)
Harming: An event *e* harms a person *S* if and only if there is a state of affairs *d* such that (i) *e* causes *d* to obtain, and (ii) *d* is a harm for *S*.
Harm: A state of affairs *d* is a harm for *S* if and only if

a there is an essential component of *d* that is a condition with respect to which *S* can be intrinsically better or worse off; and
b if *S* existed and *d* had not obtained, then *S* would be better off with respect to that condition.

As concerns clause (a), Gardner writes that although she is 'presupposing some account of well-being that specifies *respects* in which an individual can be intrinsically better or worse off', the view is nonetheless meant to be 'compatible with most substantive accounts of well-being' because (a) is 'neutral about what these respects might be' (2015, p. 434). According to Gardner, these respects might be the agent's 'health, her happiness, her interpersonal relationships, her reputation, having her desires satisfied, some combination of these, or something else altogether' (2015, p. 434). Some ideas about what these other factors might be are provided by her other arguments, which seem to require that visual capacities and intellectual abilities should also be included in the list (2015, p. 438). There are further questions to ask about how precisely (a) is meant to be understood, concerning, for instance, why it focuses only on 'essential' components of states of affairs. Gardner does not motivate this, and a natural view seems to be that *all* components of a state of affairs are essential to it. However, we shall set these questions aside in what follows.

Gardner's motivation for the first half of the antecedent of (b) – that *S* exists – is that a state of affairs, such as S*'s having poor health*, can be a harm for *S* even if *S* would not even have existed, and thus not been better off, if it had not obtained. With the clause in place, GCA yields the desired result that S*'s having poor health* is a harm for *S*, because *S* would have been better off with regard to *S*'s health if *S* had existed without poor health.[7]

GCA shares SCA's advantages in the cases already discussed. Both accounts, however, also face serious problems.

6. *Objections to SCA and GCA*

Unlike the versions of CA already criticized, SCA and GCA do not seem to undergenerate harming. Rather, they overgenerate harming in a variety of ways. They also overgenerate benefiting.

6.1. OBJECTION 1: MANY THREATS

Our first objection focuses on the following case:

Many Threats. On New Year's Eve, Blondie ties Dagwood to the railroad tracks, attaches a time bomb to his body, and hires a hundred marksmen to shoot at him. As the train is approaching, the time bomb is about to go off, and the marksmen are taking aim, Herb arrives on the scene. He has no time to avert the multiple threats to Dagwood's life, but he carries a lethal dose of morphine. Herb injects Dagwood with the morphine, causing him to experience a few moments of intense pleasure and then die painlessly of the overdose. Had Herb not acted as he did, Dagwood would have suffered a painful death a few moments later. Had he survived into the New Year, on the other hand, he would have lived happily for many more years.

Herb's action causes *Dagwood's dying on New Year's Eve* to obtain. Moreover, Dagwood would have been better off if this state of affairs had not obtained, because he would then have lived happily for many more years. Hence, SCA implies that Herb's action harms Dagwood. GCA has the same implication, because Herb's action also causes *Dagwood's not living happily for many more years* to obtain. This state of affairs appears to satisfy both clauses in GCA's criterion for a state of affairs' being a harm.

As long as we are considering *pro tanto* harming, the conclusion that Herb's action harms Dagwood may be acceptable. What is not acceptable, however, is the claim that Herb's action seriously harms Dagwood *overall*. After all, Herb's action prevents Dagwood from dying a painful death, and causes him intense pleasure; and nothing Herb could have done would have prevented Dagwood from dying on New Year's Eve. Unfortunately for SCA and GCA, they imply that Herb's action does seriously harm Dagwood overall – at least given the relation between *pro tanto* and overall harming suggested in Section 2. Given SCA and GCA, Herb's action *pro tanto* harms Dagwood to a very high degree, because the harm to Dagwood of dying on New Year's Eve is very great. The degree to which Herb's action *pro tanto* benefits Dagwood, given SCA and GCA, by causing *Dagwood's feeling some pleasure and dying painlessly* to obtain, is much lower. Summing the degrees to which Herb's action *pro tanto* harms and *pro tanto* benefits

Dagwood given SCA and GCA therefore yields the result that it seriously harms Dagwood overall. Again, this is an unacceptable conclusion.[8]

6.2. OBJECTION 2: MANY FRIENDS

In *Many Threats*, the nearest possible world in which Dagwood does not die on New Year's Eve is very far away. SCA and GCA also overgenerate harming in cases in which, intuitively, what is modally far away is rather that the relevant person is not benefited. Consider the following:

Many Friends. Batman is sick with a disease which, if left untreated, would cause him to occupy a hedonic level of 2. Fortunately, help is near: one hundred Batman fans are standing in line to cure his disease. Each of the first ninety-nine people has a pill which, if fed to Batman, would cause him to occupy a hedonic level of 5. However, the last person in line has a pill that also tastes like strawberries. If that pill were given to Batman, it would cause him to occupy a hedonic level of 8. However, only one person gets to give him a pill. The first person in line does so. No alternative available to her would have resulted in Batman's occupying a hedonic level higher than 5. If she had not given Batman a pill, the second person would have (and so Batman would have occupied a hedonic level of 5 even if the first person had not acted as she did); if neither the first nor the second had done so, the third person would have; and so on.

The first person's giving Batman the pill causes *Batman's occupying hedonic level 5* to obtain. If that state of affairs had not obtained, Batman would have been better off, because in the nearest world in which it does not obtain, the last person in line gives Batman the pill. SCA and GCA thus both yield that the first person's action harms Batman. It seems clear, however, that it does not harm him – not even *pro tanto*. Indeed, especially from the point of view of a causal theory of harming, the person's action appears to be entirely beneficial to Batman, because it causes his disease to be cured.

6.3. OBJECTION 3: MORE GOLF

As indicated in Section 5, SCA and GCA are in a better position than CCA to handle *Golf Clubs*, because Batman's decision can be said to merely allow, rather than cause, *Robin's lacking golf clubs* to obtain. However, a slight modification to the case reveals this advantage to be a minor one:

More Golf. Batman has bought two golf clubs, each of which is guaranteed to give its owner 10 units of pleasure. He can either give Robin both clubs, give him exactly one of them, or give him neither. Batman gives Robin exactly one of the clubs, whereby Robin receives exactly 10 units of pleasure. If Batman had not done so, he would have given Robin both clubs, whereby Robin would have received 20 units of pleasure.[9]

Clearly, Batman's action causes, and does not merely allow, *Robin's owning exactly one golf club* to obtain. Moreover, Robin would have owned

two golf clubs (and would have thereby been better off) if this state of affairs had not obtained. Hence, SCA yields that Batman's action harms Robin. Surely, however, if Batman's decision does not harm Robin in *Golf Clubs*, his action does not harm Robin in *More Golf* either.

As for GCA, *Robin's owning exactly one golf club* might not satisfy condition (i) – perhaps it involves no factor with respect to which Robin can be intrinsically better or worse off. But *Robin's experiencing exactly 10 units of pleasure* clearly does. Furthermore, Batman's action causes also this state of affairs to obtain, and Robin would have been hedonically better off if it had not obtained. On GCA, too, then, Batman's action harms Robin.

6.4. OBJECTION 4: MORE TEAR GAS

Not only do SCA and GCA face a variant of the failure to benefit problem, they also face a variant of the preemption problem. Consider this variation on *Tear Gas*:

More Tear Gas. The Joker, who is very determined to hurt Batman, sprays tear gas in exactly one of Batman's eyes. He does not have enough tear gas to spray it in both of Batman's eyes. Riddler, equipped with his own can of tear gas, is tempted to follow the Joker's noxious example, but eventually decides that enough is enough. Hence, if Batman had not had tear gas in exactly one eye, that would have been because Riddler would have sprayed additional tear gas in Batman's other eye (whereby Batman would have had tear gas in both eyes). If the Joker had not sprayed tear gas Riddler would have left Batman alone (whereby Batman would not have had tear gas in any eye).

In contrast to the *Tear Gas* problem for CCA, the problem for SCA and GCA here is not that they imply that the Joker's action is harmless. They do yield that the Joker's action *pro tanto* harms Batman – for instance, it arguably causes *Batman's having tear gas in at least one eye* to obtain, and Batman would have been relevantly better off if this state of affairs had not obtained. Instead, the problem is that – assuming a parallel treatment of harming and benefiting (Section 2) – SCA and GCA also imply, implausibly, that the Joker's action *pro tanto* benefits Batman. For the Joker's action also causes, for example, *Batman's having tear gas in exactly one eye* to obtain, and Batman would have had tear gas in both of his eyes (and thus been relevantly worse off) if this state of affairs had not obtained. That the Joker's action *pro tanto* benefits Batman might have been a defensible implication if, as in the original *Tear Gas*, one of the alternative actions available to the Joker would have left Batman even worse off. In *More Tear Gas*, however, there is no such alternative.

7. *Contrastive causal-counterfactual accounts*

The objections to SCA and GCA just presented can perhaps be met by incorporating the popular idea that causation is a *contrastive* phenomenon in the account of harming. On this approach, an event *e* does not simply cause an effect *d* to obtain, full stop; instead, an event *e rather than some contrast event e** causes an effect *d rather than some contrast effect d** to obtain. The relevant contrasts, *e** and *d**, are usually taken to be determined by context.[10]

In accordance with this idea, Thomas D. Bontly (2016) has proposed a view that can be translated into the language of CA as follows:

Bontly's causal-counterfactual account (BCA)
Harming: An event *e* harms a person *S* if and only if there is a state of affairs *d* and a contrast state of affairs *d** such that (i) *e* rather than a contrast event *e** causes *d* rather than *d** to obtain, and (ii) it is a harm for *S* that *d* rather than *d** obtains.
Harm: It is a harm for *S* that *d* rather than *d** obtains if and only if *d* leaves *S* worse off than *d** would have done.[11]

Robert Northcott's account (Northcott 2015) is similar to Bontly's, but makes harming itself contrastive: what does the harming is an event *e rather than* some contrast event *e**.[12] While we are not sure exactly how best to interpret Northcott, the following account is hopefully close enough (and is in any case worth considering)[13]:

Northcott's causal-counterfactual account (NCA)
Harming: An event *e* rather than contrast event *e** harms a person *S* if and only if there is a state of affairs *d* and a contrast state of affairs *d** such that (i) *e* rather than *e** causes *d* rather than *d** to obtain, and (ii) it is a harm for *S* that *d* rather than *d** obtains.
Harm: It is a harm for *S* that *d* rather than *d** obtains if and only if *d* leaves *S* worse off than *d** would have done.

As an example of how BCA and NCA handle cases that are problematic for CCA, consider again *Tear Gas*. The Joker's action (rather than the action of leaving Batman alone) apparently causes *Batman's being in pain* rather than *Batman's feeling fine* to obtain; and the former state of affairs leaves Batman worse off than the latter would have done. Hence, BCA gets this case right. NCA yields a more nuanced verdict than do BCA and other views. While NCA does not imply that the Joker's action harms Batman full stop, it arguably implies both (a) that spraying tear gas in exactly one of Batman's eyes rather than spraying tear gas in both of Batman's eyes does not harm Batman and (b) that spraying tear gas in exactly one of Batman's eyes rather than leaving him alone does harm Batman. Both (a) and (b) seem correct.

Moreover, BCA and NCA seem to handle *Many Threats* and *Many Friends* better than SCA and GCA. In *Many Threats*, the obvious contrast event to Herb's action is his refraining from giving Dagwood the morphine. And it appears false that Herb's giving Dagwood the morphine, *rather than* his refraining from doing so, causes *Dagwood's dying on New Year's Eve*, rather than *Dagwood's not dying on New Year's Eve* (or *Dagwood's living happily for many more years*), to obtain. Dagwood would for certain have died on New Year's Eve even if Herb had refrained from giving him the morphine. More generally, for any action *a*, available to Herb in the situation and incompatible with giving Dagwood the morphine, it seems false that Herb's giving Dagwood the morphine, rather than doing *a*, causes *Dagwood's dying on New Year's Eve*, rather than *Dagwood's not dying on New Year's Eve*, to obtain. Arguably, therefore, BCA and NCA avoid the implication that Herb's action harms Dagwood, even *pro tanto*.

The situation is analogous in *Many Friends*. There is no action *a*, available to the first person in line, such that her giving Batman a pill, rather than doing *a*, causes *Batman's occupying hedonic level 5* rather than *Batman's occupying a hedonic level higher than 5* to obtain. Hence, the first person's action does not harm Batman, according to BCA and NCA.

At least at first glance, BCA and NCA also seem to handle *More Tear Gas* better than SCA and GCA. In particular, it is false that the Joker's spraying tear gas in exactly one of Batman's eyes, rather than leaving him alone, causes *Batman's having tear gas in exactly one eye*, rather than *Batman's having tear gas in both eyes*, to obtain. Hence, BCA and NCA do not yield that the Joker's action benefits Batman for that reason.

As for *More Golf*, Batman's giving Robin exactly one golf club, rather than giving him two clubs, causes *Robin's owning exactly one golf club*, rather than *Robin's owning two golf clubs*, to obtain. Hence, BCA implies that Batman's action harms Robin, and is thus just as vulnerable to this case as SCA and GCA. NCA, on the other hand, implies instead that Batman's giving Robin exactly one golf club, *rather than* giving him two golf clubs, harms Robin – which is perhaps a less implausible implication than that Batman's action harms Robin full stop. This might be taken to show that if we want a contrastive view of harming and benefiting, we should go all the way and make harming itself contrastive, as NCA does.

However, the contrastive elements in BCA and NCA are of no help in the examples we will present in the following section, which are problematic for all the above causal-counterfactual accounts. Indeed, by making harming itself contrastive, NCA yields even more implausible results in some cases than the other causal-counterfactual accounts. In fact, on closer inspection this holds even for *More Tear Gas* and *More Golf*.

8. Objections to all the causal-counterfactual accounts

8.1. OBJECTION 1: PILLS

Consider the following case:

Pills. Barney suffers from a painful disease. On Monday, he can either take Pill A or not. On Tuesday, he will have another choice, between taking Pill B or not. Barney believes that he will be completely cured just in case he takes only Pill A, and partially cured just in case he takes both pills. Accordingly, he takes Pill A on Monday and does not take Pill B on Tuesday (although he would otherwise be indifferent between taking it and not). He is, however, misinformed about the effects of the pills. Taking only Pill A causes his disease to be merely partially cured. If he had taken both pills, he would have been completely cured. Had he not taken Pill A on Monday, on the other hand, nothing he could have done later would have produced even a partial cure.

In *Pills*, we take it, Barney's action of taking Pill A causes *Barney's not taking Pill B on Tuesday* to obtain.[14] Further, this state of affairs is such that Barney would have been better off if it had not obtained, given that in the nearest world in which it does not, Barney takes both Pill A on Monday and Pill B on Tuesday. For this reason, SCA implies that taking Pill A harms Barney, at least *pro tanto*. This seems clearly wrong, because taking Pill A guarantees a partial cure, and is a necessary condition for being completely cured.

Similar remarks apply to BCA and NCA. Taking Pill A, rather than refraining from taking it, arguably causes *Barney's not taking Pill B on Tuesday* rather than *Barney's taking Pill B on Tuesday* to obtain. The former state of affairs leaves Barney worse off than the latter would have done. Like SCA, then, BCA yields that taking Pill A harms Barney; and NCA yields that taking Pill A rather than refraining from taking it harms him.[15]

What about GCA? Presumably, *Barney's not taking Pill B on Tuesday* does not satisfy GCA's clause (a), and thus not (ii), as it involves nothing with respect to which Barney can be intrinsically better or worse off. However, we can simply add the stipulation that Pill B is mildly unpleasant. Assuming that a person can be intrinsically better or worse off with respect to her hedonic levels, *Barney's not taking any mildly unpleasant pill on Tuesday* satisfies (a). Because Barney's taking Pill A causes this state of affairs to obtain, and he would have been hedonically better off if he had existed and it did not obtain – he would have been completely cured and thereby happier – GCA implies that taking Pill A harms him. But the added stipulation makes it even clearer that this action does not harm Barney. In addition to guaranteeing a partial cure, and being necessary for Barney to be completely cured, it also causes some mild unpleasantness not to occur.

Our claim, in both versions of the case, that Barney would have been better off if the relevant state of affairs had not obtained, relies on the

assumption that if Barney had taken Pill B on Tuesday, he would still have taken Pill A on Monday. An alternative view is that the nearest possible world where Barney takes Pill B on Tuesday is a world in which he does not take Pill A on Monday. However, this 'backtracking' claim seems highly implausible. For example, it absurdly falsifies the statement, uttered on Tuesday morning, that if Barney were to take Pill B later in the day (contrary to his motivations), he would be completely cured. In any event, the backtracking strategy will not work against the next objection.[16]

8.2. OBJECTION 2: STONE

Our second counterexample targeting all the causal-counterfactual theories is as follows:

Stone. Catwoman is offered to throw a stone at a window. It is easy to hit the window, but more difficult to do so without breaking it. Catwoman will experience 20 units of pleasure if she hits the window without breaking it. If it breaks, she will experience 10 units of pleasure. If she declines to throw the stone, she will experience no pleasure. She throws the stone at the window, which just barely breaks. Had she used only slightly less force, or hit a slightly less fragile part of the window, it would have stayed intact. Likewise, of course, if she had declined to throw the stone.

In this case, Catwoman's throwing the stone at the window (rather than her refraining from doing so) causes *the window's breaking* (rather than *the window's not breaking*) to obtain. In the nearest world in which the window does not break, Catwoman throws the stone at the window without breaking it and is thus better off. (We assume that Catwoman knows about the details of the case and is motivated to maximize her pleasure. Given this, the nearest world in which the window does not break is surely not one where she does not throw the stone at all, but one in which she throws it with slightly less force, or hits a slightly less fragile part of the window.) For this reason, SCA and BCA both entail that throwing the stone at the window harms Catwoman. This is counterintuitive, because throwing the stone is guaranteed to leave her better off than not throwing it would.

As for GCA, Catwoman's action of throwing the stone at the window also causes *Catwoman's experiencing 10 units of pleasure* to obtain. Catwoman would have been hedonically better off if this state of affairs had not obtained, because she would then have experienced 20 units of pleasure. GCA thus also yields, counterintuitively, that Catwoman's throwing the stone at the window harms her.

NCA, finally, has even more implausible implications than SCA, BCA, and GCA. On NCA, it is not Catwoman's action of throwing the stone at the window, but her throwing the stone at the window *rather than refraining from doing so*, that harms her. And because her throwing the stone at the

window was guaranteed to leave her better off than her refraining from doing so, this implication is surely false.

The arguments just given assume that if *Catwoman's experiencing 10 units of pleasure* had not obtained (at time *t*), then she would (just before *t*) have hit the window without breaking it. This assumption can be denied on the basis of a very strict view of backtracking claims, according to which such claims are never (or almost never) true. According to this objection, in the nearest possible world *w* in which *Catwoman's experiencing 10 units of pleasure* does not obtain, she still breaks the window, just as she does in the actual world. Instead, in *w*, something else prevents her from experiencing exactly 10 units of pleasure. Moreover, because Catwoman breaks the window in *w*, there is no reason to think that she experiences 20 units of pleasure in *w* – rather, she might experience less than 10 units of pleasure. If that is so, there is no reason to think that causal-counterfactual theories entail that Catwoman's throwing the stone harms her.

Such a strict view of backtracking claims strikes us as *too* strict.[17] However, the objection can be handled without entering into that controversy. For notice that the most promising objection to our earlier argument, based on *Pills* (Section 8.1), presupposes a very *permissive* view of backtracking, on which such claims are quite often true. That objection was based on the idea that, contrary to what our argument assumed, the nearest possible world where Barney takes Pill B on Tuesday (as he actually does not) is a world in which he does not take Pill A on Monday (as he actually does). Given a very strict view of backtracking claims, then, that objection to *Pills* is certainly unsuccessful, because it requires much more backtracking than *Stone* does. In other words, there seems to be no view of backtracking that vindicates *both* objections to *Pills* and *Stone*.

8.3. OBJECTION 3: A CLOSER LOOK AT MORE TEAR GAS AND MORE GOLF

Recall this case (Section 6.4):

More Tear Gas. The Joker, who is very determined to hurt Batman, sprays tear gas in exactly one of Batman's eyes. He does not have enough tear gas to spray it in both of Batman's eyes. Riddler, equipped with his own can of tear gas, is tempted to follow the Joker's noxious example, but eventually decides that enough is enough. Hence, if Batman had not had tear gas in exactly one eye, that would have been because Riddler would have sprayed additional tear gas in Batman's other eye (whereby Batman would have had tear gas in both eyes). If the Joker had not sprayed tear gas Riddler would have left Batman alone (whereby Batman would not have had tear gas in any eye).

As noted in Section 7, BCA and NCA seem at first glance to be better equipped than SCA and GCA to handle *More Tear Gas*. On closer inspection, however, *More Tear Gas* is seriously problematic for BCA and NCA as well. The Joker's action of spraying tear gas in exactly one of Batman's

eyes, rather than leaving him alone, causes *Batman's having tear gas in an odd number of eyes*, rather than *Batman's having tear gas in an even number of eyes*, to obtain (where zero counts as an even number). And Batman would have been worse off if the latter state of affairs had obtained (because he would then have had tear gas in both eyes). Just like SCA and GCA, then, BCA yields, after all, that the Joker's spraying tear gas in exactly one of Batman's eyes benefits Batman. While this is implausible enough, the situation is even worse for NCA. For NCA implies, absurdly, that the Joker's spraying tear gas in exactly one of Batman's eyes, *rather than leaving him alone*, benefits Batman.[18]

A similar point can be made about *More Golf* (Section 6.3). As remarked in Section 7, NCA avoids the implication – common to SCA, GCA, and BCA – that Batman's giving Robin exactly one golf club harms Robin full stop. However, NCA has an even more implausible implication. For Batman's giving Robin exactly one club, rather giving him no clubs, causes *Robin's owning an odd number of golf clubs*, rather than *Robin's owning an even number of golf clubs*, to obtain. The former state of affairs leaves Robin worse off than the latter would have done (because he would then have owned two clubs). Hence, NCA implies, absurdly, that Batman's giving Robin exactly one club, *rather than giving him no clubs*, harms Robin, at least *pro tanto*.

9. Concluding remarks

In this paper, we have sought to show that no matter whether clause (ii) of CA is specified in terms of intrinsic badness, in temporal terms, or in counterfactual terms, the resulting theory is extensionally inadequate. The problem with these views, we have seen, is not just that they have counterintuitive implications in some specific case or other; rather, their implications for a great *many* cases (and cases of quite different kinds) are counterintuitive. Thus, even those who remain unconvinced by our arguments should acknowledge that adherents of CA face a challenge: either to formulate clause (ii) in an alternative way so that CA avoids these problems, or to argue that our arguments against extant versions fail for other reasons.

Moreover, in addition to posing problems for various specific versions of CA, we also take our arguments to provide strong (though of course not conclusive) reasons to think that the idea that underlies the account – that harming should be understood as causing harm – is bound to fail as well. In our view, a natural thought, and a good point of departure for any analysis of harming, is that harming is very closely connected to *negatively influencing someone's well-being* (Johansson and Risberg forthcoming). And the fundamental problem with CA (and, thus, with its various versions) seems to be that it departs too radically from this natural thought. In

Happiness Reduction, for instance, a plausible diagnosis for why CIBA fails is that it implies that the Joker's diminishing Batman's happiness does not harm Batman, despite the fact that it clearly influences his well-being negatively. Similar things can be said about the other counterexamples we have presented. More generally, as *Many Threats* illustrates, any version of CA faces the problem that an event (such as Herb's giving Dagwood the morphine) may fail to harm a person (such as Dagwood) despite causing a state of affairs that is intuitively a harm to them (such as *Dagwood's dying on New Year's Eve*) to obtain, as long as the action does not influence their well-being negatively (as Herb's action intuitively does not). A more plausible view should thus tie harming more closely to influencing someone's well-being negatively than CA does, and we suspect that the best way to do so is to abandon CA entirely.

One possible strategy in criticizing the arguments we have given is to insist that case-based intuitions about harming of the kind we rely on do not have the evidential weight we ascribe to them. However, even if this methodological approach should be plausible in its own right, it is not a promising one for adherents of CA to adopt – especially because the most serious problems for its main competitor, CCA, also have to do with the fact that it violates intuitions about harming in particular cases (such as the ones in Section 1). Thus, if such intuitions are taken to be unreliable, adherents of CA are deprived of the most important objections to its main rival. And because CCA appears to fare better than at least many versions of CA with regard to other theoretical virtues, such as simplicity, elegance, and intrinsic intuitive appeal, this methodological approach is of very little help to CA.

Finally, if our arguments against CA are sound, they have consequences for several other debates in which questions of harming and benefiting are relevant. One such debate comes from population ethics. Several versions of CA (including CIBA and the causal-counterfactual views) entail that one can harm a person by causing her to come into existence. For this reason, some of CA's proponents have taken CA to provide a harm-based solution to the so-called *non-identity problem* (e.g. Harman 2004, 2009; Gardner 2015, 2019a; Bontly 2016). Roughly, this problem concerns whether and why it is morally wrong to create a person, S, who would occupy a low but positive well-being level, rather than another person, S^*, who would occupy a much higher level. The view that we can harm S by creating her is not, we think, particularly attractive in its own right – especially if S's life would not contain any intrinsic bads (e.g. Bradley 2012, p. 406; Johansson and Risberg 2018, p. 738; cf. Carlson and Johansson 2019). But it might nonetheless be viewed as an indirect advantage of CA that it helps to explain the intuitive *wrongness* of creating S rather than S^* in cases such as these. This explanation is obviously unsuccessful, however, if CA is false. Moreover, because CCA entails that creating S, instead of S^*, *cannot* harm

S, the prospects for a harm-based solution to the non-identity problem might be looking dim.

Another debate to which our conclusions are relevant concerns the normative significance of harming. It is intuitive that we have moral reasons not to harm others, and equally intuitive that we have prudential reasons not to harm ourselves. CCA struggles to accommodate these claims, however, in part due to the preemption and failure to benefit problems and related issues (e.g. Bradley 2012; Carlson 2019, 2020; Feit 2019; Carlson et al. 2021; Johansson and Risberg forthcoming). As causal theories of harming promise to avoid those problems, they also promise to do better in accommodating the normative significance of harming. However, if causal theories are rejected – as we have suggested that they should be – then the normative significance of harming obviously cannot be vindicated by appeal to such theories. Whether some other theory of harming can be used to provide such a vindication remains, we think, to be seen.[19]

Department of Philosophy
Uppsala University, Sweden

NOTES

[1] CCA (as well as other related views of harm) faces several further problems that we lack the space to mention here but discuss elsewhere (Carlson 2019, 2020; Johansson and Risberg 2020, forthcoming; Carlson et al. 2021).

[2] That something like this relation holds between *pro tanto* and overall harming, and between *pro tanto* and overall benefiting, is widely assumed in the literature (e.g. Bradley 2012, pp. 393–394).

[3] If causation or causal contribution comes in degrees, an alternative possibility is to take the degrees to which *e pro tanto* harms *S* to depend both on the degrees of the harms it causes for *S* and on the degrees to which it causes, or causally contributes to, those harms. We shall set this possibility aside, however, because we shall only consider cases in which the degree of causation, or causal contribution, involved is intuitively maximal (or at least very high).

[4] We assume that Batman does not have to exist after *t* for this state of affairs to obtain.

[5] This case originates with Norcross (2005, pp. 165–166).

[6] According to Rabenberg (2015, p. 21), preventing *S* from becoming better off than before harms *S* only if *S* begins at a low well-being level (as Bamm-Bamm does in *Hampered Recovery*). This seems to rule out a parallel treatment of harming and benefiting (refer to Section 2). For the parallel claim about benefiting is that preventing *S* from becoming *worse* off than before benefits *S* only if *S* begins at a *high* well-being level. That cannot be right.

[7] Both (a) and (b) are apparently motivated by Gardner's view that causation requires counterfactual 'backtracking' (2015, p. 434; 2019b, p. 904). Refer further to note 16 and Section 8.2.

Pacific Philosophical Quarterly published by University of Southern California and John Wiley & Sons Ltd.

[8] In her discussion of a relevantly similar case, Gardner (2017) suggests that causal theorists of harming can accommodate our moral judgments about the relevant action by making further assumptions about the strength of our moral reasons against harming. In particular, she stipulates that an action *redundantly* harms a person just in case it causes the person a harm that she would have suffered even if the action had not been performed, and suggests that our moral reasons against redundantly harming people are weaker than our moral reasons against non-redundantly harming people (at least other things equal) (2017, pp. 80–81, 86). Our objection here, however, does not focus on whether causal theories support implausible moral conclusions, but directly on their implications about harming.

[9] This case is taken, with minor modifications, from Johansson and Risberg (2020). Unlike our other counterexamples to causal-counterfactual accounts, *More Golf* is a counterexample to CCA as well.

[10] We assume that if e is an action, then e^* is at least typically another action that the agent could have performed instead of e.

[11] While Bontly's own formulation focuses on actions rather than on events more generally, this makes no difference for present purposes, as we shall only focus on actions in what follows.

[12] For a non-causal view along similar lines, refer to Norcross (2005). Refer also to note 18.

[13] One reason why Northcott is difficult to interpret is that he is unclear about what kinds of events or states of affairs that can be effects in the relevant instances of causation. He states that a relevant effect must be 'an actual level of well-being' (2015, p. 152), but then goes on to give examples of harming in which this is clearly not the case (2015, p. 157).

[14] It may seem more natural to cite Barney's beliefs and desires, rather than his taking Pill A, as causes of his not taking Pill B. However, we are only claiming that his taking Pill A is *a* cause of his not taking Pill B, not that it is the most salient cause. Note also that it is natural to cite Barney's taking Pill A as a cause of his believing, on Tuesday, that he has taken Pill A. If this belief is a cause of his not taking Pill B, it follows, at least if causation is transitive, that his taking Pill A is a cause of his not taking Pill B.

[15] This argument focuses on a particular contrast event – refraining from taking Pill A. As an anonymous reviewer has pointed out, there are various other contrast events that do not give rise to any similar problems for BCA and NCA. One example is the contrast event of Barney's killing himself. Clearly, taking Pill A, rather than killing himself, does not cause *Barney's not taking Pill B on Tuesday* rather than *Barney's taking Pill B on Tuesday* to obtain. Equally clearly, however, none of this affects the fact that BCA and NCA have implausible implications with regard to the contrast event that is the focus of our argument. Nor is our choice of contrast event in any way unnatural or illegitimate. Analogous remarks apply to our other counterexamples to BCA and NCA.

[16] The backtracking strategy is congenial with the view presented in Gardner (2019b). Gardner suggests that CA should be combined with a view of causation on which c causes e only if c would not have occurred if e had not occurred. Another response to *Pills* is to appeal to this view of causation to deny our claim that Barney's taking Pill A on Monday causes him not to take Pill B on Tuesday, on the grounds that the relevant backtracking condition is in this case not satisfied. This move does not help with all the cases that we have discussed, however, because the relevant condition is clearly satisfied in at least some of them (including *Many Threats*, *Many Friends*, and *More Golf*).Moreover, Gardner's backtracking view of causation creates a further problem for her own version of CA, GCA. Suppose that Wilma feels intense pain and that this event causes state of affairs d = *Wilma's having a mildly pleasant memory of the pain* to obtain (maybe the pleasantness is due to her finding the earlier pain phenomenologically interesting). d essentially involves a hedonic component, and thus, a condition with respect to which Wilma can be intrinsically better or worse off. On the backtracking view, if (Wilma had existed and) d had not obtained, then she would have been hedonically better off (because she would not have had the pain). Hence, on GCA, one thing in virtue of which Wilma's pain harms her is its causing her to have a mildly pleasant memory of it. That seems clearly wrong. (Note that it will not help

to restrict GCA to events that, unlike experiences of pain, do not harm the person *intrinsically,* for the same problem arises if we replace Wilma's pain in the example with, say, a painless event that deprives her of a lot of pleasure.) Refer also to Johansson and Risberg (2018).

[17] The strict view is also in conflict with the supposition in *Ouch* (Section 1) that Wilma would not have felt any pain if she had not said 'ouch'. Thus, the strict view undermines one of the reasons to prefer CA to CCA.

[18] Note that the same can be said about the original *Tear Gas.* Note also that it is irrelevant to BCA and NCA that the closest possible world in which Batman has tear gas in an even number of eyes (which, again, is a world in which he has tear gas in both eyes) is a world in which the relevant contrast event – leaving Batman alone – does not occur. More generally, it is irrelevant to BCA and NCA how well off the person is in the closest possible world in which the contrast event (e^* in BCA and NCA) occurs. What matters is, instead, how well off the person is in the closest possible world in which the contrast *effect* (state of affairs d^* in BCA and NCA) obtains. The contrastive view in Norcross (2005) differs from BCA and NCA in this regard – unsurprisingly, as it is not a version of CA but of CCA.

[19] For very helpful comments on earlier versions, we are grateful to Justin Klocksiem, Charlotte Unruh, participants in the Higher Seminar in Practical Philosophy at Uppsala University, and several anonymous referees. Erik Carlson's and Jens Johansson's work for this paper was supported by Grant 2018-01361 from Vetenskapsrådet. Carlson also received support from Vetenskapsrådet's Grant 2016-01531. Olle Risberg's work was supported by Grant 2020-01955 from Vetenskapsrådet.

REFERENCES

Bontly, T. (2016). 'Causes, Contrasts, and the Non-Identity Problem,' *Philosophical Studies* 173, pp. 1233–1251.

Boonin, D. (2014). *The Non-Identity Problem and the Ethics of Future People*. New York: Oxford University Press.

Bradley, B. (2004). 'When Is Death Bad for the One Who Dies?' *Noûs* 38, pp. 1–28.

Bradley, B. (2009). *Well-Being and Death*. New York: Oxford University Press.

Bradley, B. (2012). 'Doing Away with Harm,' *Philosophy and Phenomenological Research* 85, pp. 390–412.

Carlson, E. (2019). 'More Problems for the Counterfactual Comparative Account of Harm and Benefit,' *Ethical Theory and Moral Practice* 22, pp. 795–807.

Carlson, E. (2020). 'Reply to Klocksiem on the Counterfactual Comparative Account of Harm,' *Ethical Theory and Moral Practice* 23, pp. 407–413.

Carlson, E. and Johansson, J. (2018). 'Well-Being without Being? A Reply to Feit,' *Utilitas* 30, pp. 198–208.

Carlson, E. and Johansson, J. (2019). 'Bontly on Harm and the Non-Identity Problem,' *Utilitas* 31, pp. 477–487.

Carlson, E., Johansson, J. and Risberg, O. (2021). 'Well-Being Counterfactualist Accounts of Harm and Benefit,' *Australasian Journal of Philosophy* 99, pp. 164–174.

Conee, E. (2006). 'Dispositions toward Counterfactuals in Ethics,' in K. McDaniel et al. (eds) *The Good, the Right, Life and Death: Essays in Honor of Fred Feldman*. Aldershot: Ashgate, pp. 173–188.

Feit, N. (2002). 'The Time of Death's Misfortune,' *Noûs* 36, pp. 359–383.

Feit, N. (2015). 'Plural Harm,' *Philosophy and Phenomenological Research* 90, pp. 361–388.

Feit, N. (2016). 'Comparative Harm, Creation and Death,' *Utilitas* 28, pp. 136–163.

Feit, N. (2019). 'Harming by Failing to Benefit,' *Ethical Theory and Moral Practice* 22, pp. 809–823.

Foddy, B. (2014). 'In Defense of a Temporal Account of Harm and Benefit,' *American Philosophical Quarterly* 51, pp. 155–165.
Gardner, M. (2015). 'A Harm-Based Solution to the Non-Identity Problem,' *Ergo: An Open Access Journal of Philosophy* 2, pp. 427–444.
Gardner, M. (2016). 'Beneficence and Procreation,' *Philosophical Studies* 173, pp. 321–336.
Gardner, M. (2017). 'On the Strength of the Reason against Harming,' *Journal of Moral Philosophy* 14, pp. 73–87.
Gardner, M. (2019a). 'David Boonin on the Non-Identity Problem: Rejecting the Second Premise,' *Law, Ethics and Philosophy* 7, pp. 29–47.
Gardner, M. (2019b). 'When Good Things Happen to Harmed People,' *Ethical Theory and Moral Practice* 22, pp. 893–908.
Hanna, N. (2016). 'Harm: Omission, Preemption, Freedom,' *Philosophy and Phenomenological Research* 93, pp. 251–273.
Hanser, M. (2009). 'Harming and Procreating,' in M. Roberts and D. Wasserman (eds) *Harming Future Persons*. Dordrecht: Springer, pp. 179–199.
Hanser, M. (2019). 'Understanding Harm and Its Moral Significance,' *Ethical Theory and Moral Practice* 22, pp. 853–870.
Harman, E. (2004). 'Can we Harm and Benefit in Creating?' *Philosophical Perspectives* 18, pp. 89–113.
Harman, E. (2009). 'Harming as Causing Harm,' in M. Roberts and D. Wasserman (eds) *Harming Future Persons*. Dordrecht: Springer, pp. 137–154.
Johansson, J. and Risberg, O. (2018). 'The Problem of Justified Harm: A Reply to Gardner,' *Ethical Theory and Moral Practice* 21, pp. 735–742.
Johansson, J. and Risberg, O. (2019). 'The Preemption Problem,' *Philosophical Studies* 176, pp. 351–365.
Johansson, J. and Risberg, O. (2020). 'Harming and Failing to Benefit: A Reply to Purves,' *Philosophical Studies* 177, pp. 1539–1548.
Johansson, J. and Risberg, O. (forthcoming). 'A Simple Analysis of Harm,' *Ergo: An Open Access Journal of Philosophy*.
Klocksiem, J. (2012). 'A Defense of the Counterfactual Comparative Account of Harm,' *American Philosophical Quarterly* 49, pp. 285–300.
Norcross, A. (2005). 'Harming in Context,' *Philosophical Studies* 12, pp. 149–173.
Northcott, R. (2015). 'Harm and Causation,' *Utilitas* 27, pp. 147–164.
Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
Perry, S. R. (2003). 'Harm, History, and Counterfactuals,' *San Diego Law Review* 40, pp. 1283–1314.
Purves, D. (2019). 'Harming as Making Worse off,' *Philosophical Studies* 176, pp. 2629–2656.
Rabenberg, M. (2015). 'Harm,' *Journal of Ethics and Social Philosophy* 8, pp. 1–32.
Shiffrin, S. (1999). 'Wrongful Life, Procreative Responsibility, and the Significance of Harm,' *Legal Theory* 5, pp. 117–148.
Shiffrin, S. (2012). 'Harm and Its Moral Significance,' *Legal Theory* 18, pp. 357–398.
Smuts, A. (2012). 'Less Good but Not Bad: In Defense of Epicureanism about Death,' *Pacific Philosophical Quarterly* 93, pp. 197–227.
Suits, D. B. (2001). 'Why Death Is Not Bad for the One Who Died,' *American Philosophical Quarterly* 38, pp. 69–84.
Thomson, J. J. (2011). 'More on the Metaphysics of Harm,' *Philosophy and Phenomenological Research* 82, pp. 436–458.
Timmerman, T. (2019). 'A Dilemma for Epicureanism,' *Philosophical Studies* 176, pp. 241–257.
Velleman, J. D. (2008). 'The Identity Problem,' *Philosophy & Public Affairs* 36, pp. 221–244.