Ergo AN OPEN ACCESS JOURNAL OF PHILOSOPHY

A Simple Analysis of Harm

JENS JOHANSSON
Uppsala University

OLLE RISBERG
Uppsala University

In this paper, we present and defend an analysis of harm that we call the *Negative Influence on Well-Being Account* (NIWA). We argue that NIWA has a number of significant advantages compared to its two main rivals, the *Counterfactual Comparative Account* (CCA) and the *Causal Account* (CA), and that it also helps explain why those views go wrong. In addition, we defend NIWA against a class of likely objections, and consider its implications for several questions about harm and its role in normative theorizing.

Keywords: harm; benefit; well-being; the counterfactual comparative account of harm; the causal account of harm

1. Introduction

Philosophers have recently shown a great deal of interest in what it is for an event, such as an action, to *harm* someone.

This is a natural interest, especially in view of the central role that harm plays in much normative thinking. It is essential to various deontological moral principles, such as the doctrine of doing and allowing harm. And one need not be a deontologist to hold that if an action would harm people (and some alternative action would not), then there is some moral reason against the action—and some prudential reason for each potential victim to prevent it, if possible. In order to understand what such normative claims amount to, and to apply them correctly in specific cases, it seems important to find out what harm is.

Where to start? An attractive first thought—and one that many writers seem to share—is that harming an individual somehow involves *adversely affecting her*

Contact: Jens Johansson <jens.johansson@filosofi.uu.se>
Olle Risberg <olle.risberg@filosofi.uu.se>

https://doi.org/10.3998/ergo.2275

well-being, or (in other words) *having a negative influence or impact* on her well-being. (In contemporary vocabulary, a harmful event is a negative "influencer," with respect to well-being.) Various accounts of the nature of harm can plausibly be regarded as attempts to do justice to this attractive first thought. This holds, in particular, for the two most prominent accounts. One of these is the *Counterfactual Comparative Account* (CCA):

> CCA What it is for an event *e* to harm an individual *S* is for it to be the case that *S* would have been better off if *e* had not occurred. (E.g., Boonin 2014; Bradley 2004; 2009; Feit 2002; 2015; 2016; 2019; Klocksiem 2012; Parfit 1984: 69; Timmerman 2019.)

The other one is the *Causal Account* (CA):

> CA What it is for an event *e* to harm an individual *S* is for *e* to cause *S* to be in a bad state. (E.g., Bontly 2016; Gardner 2015; 2016; 2019a; 2019b; Hanser 2009; 2019; Harman 2004; 2009; Northcott 2015; Rabenberg 2015; Shiffrin 1999; 2012; Smuts 2012; Suits 2001; Thomson 2011; Velleman 2008.)

Different versions of CA construe "bad state" differently—for instance, as a state that is *intrinsically* bad for *S*, or a state in which *S* is worse off than before, or a state such that *S* would have been better off if she had not been in it.

Which of CCA and CA is true? Probably neither. Both of them (in their various versions) have been subjected to severe harm, in the form of seemingly lethal counterexamples and other types of criticism. Other accounts have tended to fare even worse. These problems have led some writers to some rather drastic measures. In particular, Ben Bradley contends that the grave failures of existing accounts reveal the notion of harm to be too messy and ill-behaved for serious ethical theorizing; harm, he proclaims, deserves to "go the way of phlogiston" (Bradley 2012: 411).

We propose a different approach. A good analysis of harm, we suggest, is right under our noses. Recall the "attractive first thought" above: harming an individual somehow involves adversely affecting her well-being. This thought, we submit, straightforwardly delivers a plausible analysis of harm; call it the *Negative Influence on Well-Being Account* (NIWA):

> NIWA What it is for an event *e* to harm an individual *S* is for *e* to adversely affect *S*'s well-being.

As already indicated, while CCA, CA, and similar proposals might be seen as attempts to do justice to the attractive first thought, they constitute highly

problematic accounts of harm. As we shall see, this is largely due to their not taking *seriously enough* the attractive first thought. By contrast, NIWA wholeheartedly embraces it, and thereby avoids those problems. In our view, then, NIWA is substantially more than a good starting point for the discussion. Despite being possibly less illuminating than views like CCA and CA, it is nonetheless a legitimate account in its own right, and a promising *rival* to such accounts.

In this paper, we shall defend NIWA, and highlight various ways in which it affects several different controversies surrounding the nature and normative significance of harm.

The paper is organized as follows. Section 2 provides some preliminary remarks. In Section 3, we offer several kinds of support for NIWA. Among other things, we argue that NIWA survives various thought experiments that have been taken to bring down CCA and CA, and also helps to explain why these theories go wrong. We also argue that NIWA has several other virtues as well, such as accommodating the normative significance of harm and avoiding the conclusion that harm is a "contextual" matter. In Section 4 we address a class of likely objections, to the effect that NIWA is circular, or an obvious truth that everyone in the debate already accepts, or otherwise insufficiently illuminating. In Section 5, we argue that NIWA helps to expose as ill-founded Bradley's skepticism about harm and its role in ethical theory. Section 6 concludes.

2. Preliminary Notes

We begin with three sets of preliminary remarks.

First, when authors in the debate discuss some "analysis," "account," "theory," or "view" of harm, there is sometimes room to wonder what exactly is being considered. Is the relevant proposal meant to be a linguistic analysis (specifying the meaning of the word 'harm'), or a conceptual analysis (specifying the content of the concept HARM), or a metaphysical analysis (specifying the nature of the phenomenon harm), or a statement of the grounds of harm (specifying in virtue of what harm obtains), or merely a statement of necessary and sufficient conditions for harm, or perhaps something else? This is not always made clear. Similarly, NIWA can also be understood in different ways: as a linguistic analysis, a conceptual one, a metaphysical one, and so on. We are inclined to think that NIWA provides a plausible analysis of the concept HARM, and one that also reveals the nature of what the concept HARM stands for—that is, of harm.[1] For example, it is plausible that there is a conceptual connection between harm

1. Of course, this is not to say that all adequate conceptual analyses reveal the nature of that for which the relevant concepts stand. It is well known, for instance, that an adequate analysis of the concept WATER does not reveal that the nature of water can be understood in terms of H_2O.

and well-being, and this is a connection that NIWA, understood as a conceptual analysis, captures. Likewise, there seems to be little reason to deny that this conceptual analysis, if correct, reveals well-being to be part of the nature of harm. That said, not much will depend on these details in what follows. While we will sometimes point out virtues of NIWA when conceived of as a conceptual analysis, what is important to us is for the most part only that NIWA is an informative account that implies that an event's adversely affecting someone's well-being is necessary and sufficient for it to harm her.

Second, in our vocabulary, whenever an event harms someone, it is a harm to her, and vice versa. Many advocates of CA apparently wish to reserve the noun 'harm' for the bad states that, according to CA, events that harm us cause us to be in (see, e.g., Gardner 2015; 2019b). These writers would say that views like CCA, CA, and NIWA are accounts of harming only, not harm. While not much hangs on this issue (but see Section 3.1), we find our vocabulary more natural. Compare: an event surprises you just in case it is a surprise to you; an event worries you just in case it is a worry for you; an event troubles you just in case it is a trouble for you.

Third, it is widely (though not universally) held that parallel accounts should be given of harm and benefit. This idea implies that we should accept NIWA only if we should also accept the following *Positive Influence on Well-Being Account* (PIWA) of benefit:

> PIWA What it is for an event *e* to benefit an individual *S* is for *e* to positively affect *S*'s well-being.

While we will mainly focus on harm in what follows, we accept the idea that parallel accounts of harm and benefit should be given, and see PIWA as equally promising as NIWA.

3. Support for NIWA

In this section, we shall adduce various specific considerations in support of NIWA. First, however, we wish to highlight a more general point that is going to underlie much of the following discussion. It seems to us that even before looking at specific kinds of support for NIWA, everyone should have suspected that CCA and CA, unlike NIWA, are on the wrong track. For given what we have called the "attractive first thought"—that harm somehow involves affecting the subject's well-being adversely—views that move substantially away from it are bound to be problematic. And while we do not offer any analysis of the notion of adversely affecting something, we take it that our grasp of it is sufficiently

good for us to see that it cannot be captured simply in terms of the counterfactual and causal relations that CCA and CA invoke. This can be illustrated by a variety of cases outside the particular context of harm and well-being. To begin with, it seems that a stock market crash can adversely affect the economy even if it is true that, had it not occurred, some other event would have occurred which would have affected the economy at least equally adversely. Moreover, an expert's belief that the economy is in bad shape can be such that if she had lacked it, the economy would not have been in bad shape. Surely her belief need not thereby *affect* the economy adversely. Such cases suggest that the notion of adversely affecting something cannot straightforwardly be understood wholly in counterfactual terms. Other cases suggest that it cannot be understood wholly in terms of causation either. For example, it is plausible that deliberately doing something morally wrong can negatively influence one's moral worth, or that a reciter's mediocre vocal performance can negatively influence the aesthetic value of a poetry reading, even though deliberately doing something morally wrong, or the reciter's mediocre vocal performance, does not plausibly *cause* any state, event, or fact that has to do with one's moral worth, or the aesthetic value of the poetry reading—the relevant relation here is not causation, but rather something like metaphysical grounding. The relations invoked by CCA and CA, then, render them—unlike NIWA—unable to fully accommodate the attractive first thought. This already should ring an alarm bell.

We now proceed to offer four more specific kinds of support for NIWA, ones that will largely reflect this general point. The first is that NIWA has intuitively plausible implications in a wide range of cases, including ones that have been taken to constitute counterexamples to CCA and CA, and it also provides a natural diagnosis for why those views fail in those cases (§3.1). Further support for NIWA is that it allows unclear cases of harm to remain unclear (§3.2), that it preserves the normative importance of harm (§3.3), and that it nicely accounts for several distinctions between different kinds of harm (§3.4). Most of the problems for CCA and CA that we will consider have already been discussed in detail elsewhere in the literature; our aim here is not to provide an exhaustive discussion of those problems, or to argue that they are impossible to overcome, but only to show that NIWA handles them much better than its main competitors.

3.1. Extensional Adequacy

The debate about harm has primarily focused on the extensional adequacy of the accounts of harm that have been offered. The idea, as Bradley puts it, is that an analysis of harm—just like an analysis of any notion—should "fit the data" (2012: 394). It should neither overgenerate nor undergenerate harm. For instance,

he claims, an analysis of harm should not entail that killing someone is normally harmless to her.

That first test case is no match for NIWA: clearly, killing someone normally affects her (lifetime) well-being adversely. NIWA also accommodates various other paradigmatic harms, such as losing one's job, missing one's flight, and breaking one's leg. Typically, if an event belongs to any of these types, it adversely affects the subject's well-being. And insofar as it happens not to—for example, if the job that the person lost was pointless for her to keep—it is correspondingly less plausible that the event harms her. NIWA also accommodates various paradigmatic *non*-harms, such as taking a walk, drinking a glass of water, and reading a book. Most events of these types neither harm the subject nor negatively influence her well-being, and insofar as such an event does negatively impact the subject's well-being—for example, if the subject's taking a walk causes her to get a cold—it is correspondingly plausible that it does harm her.

Of course, any account of harm worth considering succeeds in many easy cases. More interestingly, however, NIWA also handles various cases that have proved problematic for its two main rivals: CCA and CA. We shall first look at some problems for CCA, and then some for CA.

The so-called *preemptive harm* problem for CCA (e.g., Bradley 2012; Hanna 2016; Norcross 2005) can be illustrated by the following case:

> *Tear Gas.* The Joker sprays tear gas in exactly one of Batman's eyes. If he had not done so, 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. CCA implies that it does not, however, since Batman would have been even worse off (and thus not better off) if the Joker had not done what he did. Despite this, it is still plausible that what the Joker actually does—spraying tear gas in exactly one of Batman's eyes—adversely affects Batman's well-being. (Compare with the stock market crash example above.)[2]

A second problem for CCA, the *preemptive benefit* problem, is a mirror image of the first one. It can be illustrated by the following case (see Johansson & Risberg 2020):

2. An anonymous referee has suggested that *Tear Gas* is not properly regarded as a *preemption* case, as it does not involve two potential causes one of which prevents the other from causing some relevant effect. Be that as it may; *Tear Gas* is still a counterexample to CCA. Similar remarks apply to *Slightly Worse Golf Clubs* below. For some relevant further discussion, see, e.g., Boonin (2014: 62–63), Feit (2015), and Johansson and Risberg (2019).

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

Intuitively, Batman's giving Robin the slightly worse golf clubs does not harm Robin. CCA is incompatible with this verdict since Robin would have been even better off if Batman had not performed that action. NIWA accords with this verdict, however, since Batman's action plausibly does not affect Robin's well-being negatively. Instead, his action arguably affects Robin's well-being *positively*, which means that PIWA (see Section 2) implies that his action benefits Robin. This is also an intuitively plausible result.

A third problem for CCA concerns events that are *indicators* of harm without being harms themselves. Here is one example (cf. Carlson, Johansson, & Risberg 2022: 422):

> *Omniscience.* Robin feels intense pain. As a result, an essentially omniscient being, *O*, immediately forms the belief that Robin feels intense pain. If *O* had not formed that belief, it would have been because Robin didn't feel intense pain.

CCA implies, counterintuitively, that *O*'s forming the belief that Robin feels intense pain harms Robin. NIWA, by contrast, plausibly implies that it does not. For, similarly to what we noted about the expert's belief about the economy above, *O*'s forming that belief plausibly does not affect Robin's well-being negatively—it is only his pain, as well as whatever led up to it, that has a negative *impact* on his well-being.

Not only does NIWA have more plausible implications than CCA in these cases—it also provides a plausible *motivation* for these implications. In support of the verdict that the Joker's action harms Batman in *Tear Gas*, it is natural to appeal to its negative impact on his well-being; in support of the verdict that Batman's action benefits, rather than harms, Robin in *Slightly Worse Golf Clubs*, it is natural to appeal to its positive impact on his well-being; and in support of the verdict that *O*'s forming the belief that Robin is in pain does not harm Robin (and also does not benefit him) in *Omniscience*, it is natural to appeal to the fact that it has no negative (or positive) impact on his well-being. All this obviously accords with NIWA.

Turning now to CA, this view (to repeat) states that for an event to harm someone is for it to cause her to be in a bad state. Several different versions of this can be obtained by specifying "bad state" in different ways. However, although most of those versions avoid at least some of the problems for CCA, they too face significant problems (for further discussion see Carlson, Johansson, & Risberg 2022).

To begin with, if a bad state is understood as a state that is *intrinsically* bad for the subject, then the resulting view, unlike NIWA, arguably fails with a variant of Bradley's first test case mentioned above: an account of harm should not entail that killing someone painlessly typically does not harm her. Painless killing normally does not cause the victim to be in an intrinsically bad state; in particular, being dead (it is widely held) is not such a state. Hence, the present version of CA counts painless killing as harmless. For the same reason, and again unlike NIWA, this version of CA also counterintuitively counts *the event of death* as harmless.[3]

On a *temporal* version of CA, an event is harmful just in case it causes the subject to be worse off than she was *before* the event occurred. Cases like the following pose problems for this view (Carlson, Johansson, & Risberg 2022: 428; cf. Rabenberg 2015: 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. As a result, his well-being level after *t* is the same as it was before *t*.

Intuitively, Betty's action harms Bamm-Bamm. This verdict accords with NIWA, since Betty's action also intuitively adversely affects Bamm-Bamm's well-being. However, since Betty's action does not cause Bamm-Bamm to be worse off than he was earlier, this verdict is incompatible with the temporal version of CA.

Perhaps the most promising versions of CA are versions of what can be called the *causal-counterfactual* view. On the simplest such version, an event harms a subject just in case it causes her to be in a state such that she would have been better off if she had not been in it (for more elaborate variants, see Bontly 2016; Gardner 2015; and Northcott 2015). While this view arguably avoids most problems for other versions of CA and for CCA, it does not avoid all of them. Consider

3. Purves (2016) claims that only CCA accommodates the harm of the event of death. In our view, however, not only NIWA but also several versions of CA, such as Molly Gardner's account (Gardner 2015), do too.

again *Slightly Worse Golf Clubs*. Batman's action causes Robin to be in the state of having merely very good golf clubs, and (we may suppose) if he had not been in that state he would have been even better off (since Batman would then have given him the even better clubs). Like CCA, then, the causal-counterfactual view has the implausible result that Batman's action harms Robin.[4]

The causal-counterfactual view also has problems that the other views avoid. Consider this case (for some others, see Carlson, Johansson, & Risberg 2022: 433):

> *Many Friends.* Batman is sick with a disease which, if left untreated, would cause him to occupy a low well-being level. 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 moderately high well-being level. 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 somewhat higher well-being level. 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 higher well-being level. If that person had not given Batman a pill, the second person would have (and so Batman would have occupied a moderately high well-being level 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.

Intuitively, the Batman fan's action of giving Batman the pill does not harm him. (Instead, intuitively, it benefits him because it cures his disease.) This verdict

4. Some versions of the causal-counterfactual view involve a *contrastive* element (see, e.g., Bontly 2016 and Northcott 2015). On the most natural such version, *e rather than some contrast event e** (as opposed to *e*, full stop) harms a person just in case *e* rather than *e** causes some state of affairs *d*, *rather than some contrast states of affairs d**, to obtain, and *d* leaves the person worse off than *d** would have done. As contrastivist views introduce several further complications, and we criticize them thoroughly elsewhere (Carlson, Johansson, & Risberg 2022; cf. Carlson & Johansson 2019), we will not spend much more time on them here. Two problems are briefly worth noting, however. The first is that cases like *Slightly Worse Golf Clubs* are even more problematic for the contrastive view just formulated than for non-contrastive versions of the causal-counterfactual view. Suppose that the extremely good set contains two golf clubs, whereas the merely very good set contains only one. Batman's giving Robin the merely very good set, *rather than giving him no clubs*, causes *d = Robin's owning an odd number of golf clubs*, rather than *d* = Robin's owning an even number of golf clubs*, to obtain (where zero counts as an even number). Suppose that the closest possible world in which *d** obtains is one in which Robin gets the excellent set with two golf clubs (that world is thus closer to the actual world than is the one in which Batman gives Robin no golf clubs). In that case, *d* leaves Robin worse off than *d** would have done. Hence, the contrastive view implies, absurdly, that Batman's giving Robin the slightly worse set of clubs, *rather than giving him no clubs*, harms Robin. The second problem is that contrastivist views struggle to accommodate the *normative significance* of harm; see further sect. 3.3, especially footnote 14.

accords with NIWA (and with PIWA), since the fan's action does not plausibly have a negative impact (but instead a positive one) on Batman's well-being. By contrast, the causal-counterfactual view implies that the Batman fan's action harms him. For it causes him to be in the state of *occupying a moderately high well-being level*, and in the nearest possible world in which Batman is not in that state, he occupies a somewhat higher well-being level instead, since in that world he is fed the pill that tastes like strawberries.[5]

There are many other possible versions of CA. However, instead of engaging in detail with those views here, we will note a general problem for all versions of CA—that harm sometimes involves non-causal grounding rather than causation. Here is one illustration:

> *C-fibers.* Batman's C-fibers are stimulated at *t*. This event non-causally makes him feel pain at *t*, but causes nothing in particular to happen.

Intuitively, the stimulation of Batman's C-fibers harms him, since it makes him feel pain. This verdict accords with NIWA, since the stimulation of his C-fibers plausibly has a (non-causal) negative impact on his well-being. (Compare with the "moral worth" and "poetry reading" examples above.) No version of CA can accommodate this judgment, however, as the stimulation of Batman's C-fibers does not cause anything in particular to happen.

In this case, the stimulation of Batman's C-fibers plausibly fully grounds his pain. But CA is also threatened by cases of *partial* grounding. In particular, an event can harm someone by being a constituent of something that is intrinsically bad for her. While many different axiologies could be used to illustrate this point, let us for simplicity assume that any compound fact of the following kind,

5. Someone who accepts a very strict view on backtracking might deny this counterfactual, and contend that in the nearest possible world in which Batman does not occupy a moderately high well-being level, the first person in line still gives Batman the pill but the pill does not work (and Batman thus remains at a low well-being level). Similar moves might be made with regard to what we say above about *Slightly Worse Golf Clubs*, in connection with the causal-counterfactual account. We find it implausible, however, to suppose that the details of the cases cannot even be spelled out in such a way as to make the relevant backtrackers true. Moreover, such a strict view of backtrackers, though apparently endorsed by Bontly (2016: 1248), is also for other reasons likely to be unattractive to advocates of the causal-counterfactual account. For instance, the strict view threatens to rule out the coherence of *Omniscience*, a case that is otherwise a reason to prefer the causal-counterfactual account (and other versions of CA) to CCA. Also, various key claims made by Gardner, perhaps the most ardent advocate of the causal-counterfactual account, rely heavily on a very *liberal* view on backtracking—and indeed, on the claim that if *c* causes *e*, then if *e* had not taken place, then *c* would not have taken place (see Gardner 2019b; and also 2015). In Carlson, Johansson, and Risberg (2022: 432–40), moreover, we provide various further counterexamples to the causal-counterfactual account, and argue that neither a strict nor a liberal view of backtracking can help to defuse all of them. See also the "Dog-bite" case in Bontly (2016: 1245). Thanks to an anonymous referee for helpful comments regarding these matters.

but no constituent of such a fact, is intrinsically bad for *S*: *its being the case both that* e *occurs and that* S *desires* e *not to occur*. Now consider:

> *Synthetic fibers.* Robin strongly desires not to wear clothes made from synthetic fibers. But he does so. His doing so causes nothing in particular to happen.

Given the axiology just presented, Robin's wearing clothes made from synthetic fibers seems to harm him, as it is a constituent of (and thus partially grounds) something that is intrinsically bad for him. No version of CA accommodates this, as CA's causal requirement is not satisfied. By contrast, NIWA yields that Robin's wearing clothes made from synthetic fibers does harm him. Since that event partially grounds something that is intrinsically bad for him, it surely adversely affects his well-being.

A possible response to these problems departs from the idea that we should distinguish between the noun '(a) harm' and the verb '(to) harm', and take harms to be states of affairs rather than events.[6] Given this distinction, CA can be formulated as the view that *for* e *to harm* S *is for* e *to cause harm to* S. (This idea contrasts with our second preliminary note in Section 2, but is still worth considering.) And a supporter of CA (thus formulated) may insist that our harm-related intuitions about *C-fibers* and *Synthetic fibers* are due to the presence of *harms*, rather than of things *that harm*. More precisely, the idea would be that even though the event of Batman's C-fibers being stimulated at *t* does not harm him, the state of affairs *Batman's C-fibers are stimulated at* t is nonetheless a harm to him (and mutatis mutandis for *Synthetic fibers*). This, one might think, suffices to accommodate our harm-related intuitions about these cases.

This response can be criticized on several grounds. To begin with, whether or not the state of affairs *Batman's C-fibers are stimulated at* t is a harm to him, it is still very plausible that the event of his C-fibers being stimulated at *t* harms him. Indeed, if anything, the assumption that the relevant state of affairs is a harm to him makes it *more*, rather than less, plausible that the corresponding event harms him. CA remains incompatible with this verdict about that event—and in fact even with the verdict that the event *either* harms Batman *or* is a harm to him—even given the proposed distinction between the noun 'harm' and the verb 'harms'.

Moreover, the response just outlined assumes that the state of affairs *Batman's C-fibers are stimulated at* t counts as a harm according to CA. And this assumption too can be questioned, since at least most versions of the view either do not or need not have that consequence. For one, the "intrinsic badness" version of the

6. Thanks to an anonymous referee for prompting us to address this response.

view obviously excludes that that state of affairs is a harm to Batman, since it is not intrinsically bad for him (even though it grounds something that is intrinsically bad for him).[7] The temporal view has the same result in at least some versions of the case; that is, those in which Batman was equally badly off before *t* (perhaps because his C-fibers were stimulated before *t* too). And the causal-counterfactual version also has that result in some versions of the case; that is, those in which it is true that if Batman's C-fibers had not been stimulated at *t*, something else would have happened which would have left him at the same well-being level. In such variations of the case, these versions of CA have the implausible result that no events involved harm Batman and no states of affairs involved are harms to him.[8]

Finally, a general problem for CA, which the formulation of CA in terms of harming and causing harm does not avoid, is that an event that causes something that is intuitively a harm (or a bad state) to someone need not also harm her. A simple revision of *Many Friends* illustrates this.[9] Suppose that the Batman fan's action of giving Batman the pill causes Batman to occupy a moderately negative well-being level; thanks to the pill, his well-being level goes up from very negative to mildly negative. As in the original version of the case, this action leaves Batman better off than he would have been if he had been given no pill at all, and equally well off as he would have been if any of the other first ninety-nine people in line had given him a pill, and somewhat worse off than he would have been if the last person in line had given him a pill (something that, again, happens only in a remote possible world). Intuitively, Batman's occupying a moderately negative well-being level is a harm to him (and a bad state). But just as in the original version of the case, it is clear that the Batman fan's action does not harm Batman.[10]

---

7. If one accepts the controversial view that Batman's pain is *identical* to the firing of his C-fibers, then the same point can instead be made with respect to whatever it is that grounds that firing (e.g., something on the microphysical level). Note also that on any plausible view of well-being, the relevant state of affairs also does not count as a harm on Gardner's (2015) version of CA, since that view implies that a state of affairs is a harm to a subject *S* only if it involves "a condition with respect to which *S* can be intrinsically better or worse off" (Gardner 2015: 434).

8. Two anonymous reviewers have suggested that adherents of CA can respond to this challenge by adopting some revised, disjunctive version of their view, such as the view that for *e* to harm *S* is for *e* to *either* cause *or* ground a state of affairs that is a harm to *S*. However, the disjunctiveness of this view surely speaks against it, and also raises the question of what (if anything) the two disjuncts—causing harm and "grounding harm"—have in common. And NIWA provides a unified answer to this question: causing harm and grounding harm are both ways in which an event can negatively influence someone's well-being level. (See further Section 3.4 for a discussion of other ways in which NIWA supports a more unified view of harm than its rivals.)

9. For a different kind of case that illustrates the same thing, see Carlson, Johansson, and Risberg (2022: 432).

10. In her discussion of a case of this sort, Harman (2009: 148–50), too, seems to end up denying that causing someone something that is a harm, or a bad state, for her is sufficient for harming her—and thus seems to end up denying CA. Harman provides a couple of other candidate sufficient

What this illustrates is that the connection between harming and causing harm is not as tight as the formulation of CA in terms of harming and causing harm takes it to be. An event may cause someone harm without also harming them, as the revised version of *Many Friends* shows; and an event may also harm someone without causing harm to them, as *C-fibers* and *Synthetic fibers* both show.[11] NIWA, by contrast, has no trouble dealing with these cases: in *C-fibers* and *Synthetic fibers*, the relevant events plausibly impact the relevant subjects' well-being negatively, and in *Many Friends* (in either version), the Batman fan's action has no such impact.

Just as NIWA has more plausible implications than CCA, then, it also has more plausible implications than CA. Moreover, just as NIWA provides plausible motivations for the intuitive verdicts that conflict with CCA, it also provides plausible motivations for the intuitive verdicts that conflict with CA. In *Hampered Recovery*, for instance, the harmfulness of Betty's action can naturally be supported by appeal to its negative impact on Bamm-Bamm's well-being, and in *Many Friends* (in either version), the harmlessness of the fan's action can be supported by appeal to its lack of such an impact.

To conclude this subsection, we will note that more generally, NIWA provides explanations both for why CA and CCA have plausible implications in many cases (and, thus, for why they might initially seem to be promising accounts of harm) and for why these views ultimately fail nonetheless. As indicated in Section 1, part of what makes CA and CCA appealing is that they provide initially promising ways of accommodating the "attractive first thought" that harming a subject somehow involves affecting her well-being negatively. In many cases, the causal and counterfactual relations invoked by CA and CCA coincide with the relation of negatively affecting someone's well-being, and in such cases the implications of these views are as plausible as those of NIWA. As we have just seen, however, there are also several cases where CA and CCA have implausible implications—and just as NIWA would have us expect, those are precisely the cases where the relevant relations do *not* coincide. In other words, while CA and CCA apparently take the attractive first thought as their point of departure, they simply depart from it too much—whereas NIWA's loyalty to it pays off.

---

conditions for harming, but provides nothing like an account of what harming *is*, and hence no account with which we can properly compare CCA, CA, or NIWA.

11. Following Thomson (2011), Gardner claims that the verb 'harm' resembles 'freeze' and 'clean' in that they are all "causal verbs" (2019b: 901). According to Gardner, 'freeze' and 'clean' mean *cause to be frozen* and *cause to be clean*, respectively, and so "linguistic parity" (2019b: 901) suggests that the verb 'harm' similarly means *cause to be harmed*. However, just as the cases discussed above suggest that 'harm' does not (*contra* Gardner) mean *cause to be harmed*, we think it is even more clear that 'freeze' and 'clean' do not mean *cause to be frozen* and *cause to be clean*, respectively. If a queen orders her servants to clean the castle and freeze some food, for instance, she may thereby cause the castle to be clean and the food to be frozen, but she surely has not herself cleaned the castle or frozen the food. Thanks here to Erik Carlson. See also Johansson and Risberg (2018).

3.2. Unclear Cases of Harm

Sometimes it is unclear whether a case involves harm or not. A further virtue of NIWA, understood as a conceptual analysis, is that it makes sense of this unclarity. We will give three examples.

Consider the following often discussed case (taken from Bradley 2012; see also Feit 2019; Hanna 2016; Johansson & Risberg 2020; Purves 2019):

> *No Golf Clubs.* Batman contemplates giving a set of golf clubs to Robin, but eventually decides to keep them for himself. If he had not decided to keep them, he would have given the clubs to Robin, which would have made Robin better off.

Does Batman's decision harm Robin? That seems unclear. Some factors—for example, that the decision prevents Robin's well-being level from increasing—support an affirmative answer, whereas other factors—for example, that the decision leaves Robin just as well off as he would have been if Batman had not acted at all—point in the other direction. In part for these reasons, the status of Batman's decision is a controversial issue in the literature, with some writers (including Feit 2019 and Hanna 2016) claiming that it is harmful, while others (including Bradley 2012 and Purves 2019) claim that it is harmless. Since each view has something going for it, perhaps the most reasonable response is to suspend judgment about whether Batman's decision harms Robin. This predicament is only to be expected if NIWA is a correct analysis of the concept HARM. For the different factors just mentioned also seem to render it unclear, and perhaps even motivate suspending judgment about, whether Batman's decision affects Robin's well-being level adversely.

Here is another example:

> *Creation.* Betty can either create Bamm-Bamm or abstain. If she does create him, he will have an overall good life, but will also be likely to get a lethal disease at the age of ten. Betty does create Bamm-Bamm. He does get the disease at the age of ten, and dies shortly thereafter.

Does Betty's action harm Bamm-Bamm? Again, this seems unclear. On the one hand, he owes his worthwhile existence to her action. On the other hand, he also owes his lethal disease, without which he would presumably have lived to old age, to her action. In this case too, these factors help explain the disagreement about whether procreative actions such as Betty's can harm the person that is created, with some (including Gardner 2015) arguing that they can while others (including Boonin 2014) argue that they cannot. Since each view, again, has something going for it, perhaps suspending judgment about this question is warranted as well. As

before, the conceptual analysis offered by NIWA makes good sense of this. For the factors that make it unclear whether Betty's action harms Bamm-Bamm also make it unclear whether it affects Bamm-Bamm's well-being adversely.

Finally, consider this case:

> *Conversation.* The Joker and Riddler are in the middle of an intense conversation. Anything Riddler does as a response to the Joker's latest remark, including saying nothing at all, is going to substantially—and to the same degree—annoy the Joker. The Riddler says nothing.

Does the Riddler's being silent harm the Joker? This too seems unclear. On the one hand, it seems to cause the Joker a certain amount of irritation; on the other hand, nothing Riddler could have done would have avoided that, and being silent also seems the most passive of his alternatives. So perhaps suspending judgment about this question is warranted as well.

This kind of case has been less discussed than the two other cases just considered. However, one may suspect that were such a discussion to be had, it would also result in disagreement, since (at least given certain natural assumptions about the details of the case) CCA implies that Riddler's being silent is harmless while most versions of CA imply that it is harmful. As before, NIWA makes sense of this unclarity, since the considerations that make it unclear whether Riddler's being silent harms the Joker also render it unclear whether the Riddler's being silent affects the Joker's well-being negatively.

Some might not agree that it is unclear whether the relevant actions are harmful in *No Golf Clubs*, *Creation*, or *Conversation*. Such a reaction, though, would likely support NIWA in a different way. For those who find it clear that the relevant action does, or does not, harm the subject in some of these cases will likely find it correspondingly clear that the action does, or does not, negatively affect the subject's well-being. The only way to argue that one of these cases poses problems for NIWA is to argue *both* that the relevant action harms the subject *and* that it does not affect her well-being negatively, or vice versa—a combination of views that is likely to strike most people as unattractive.

3.3. The Normative Significance of Harm

It seems undeniable, at least very generally speaking, that harm is prudentially and morally significant. This accords with NIWA, for it seems just as undeniable that whether an action affects people's well-being is, in general, prudentially and morally significant. Now, different theorists disagree on exactly *how* normatively significant harm is—for instance, is it quite as central to morality as many deontol-

ogists think? (For discussion, see, e.g., Bradley 2012; Feit 2019; Hanna 2016; Hanser 2019; Shiffrin 2012.) This is a difficult issue, and an analysis of harm should not be expected to have very much to say about it. Indeed, a virtue of NIWA is that is neutral on this issue, and makes sense of the different sides of the debate. It is easy to see how different normative theorists can rationally disagree on the precise extent of the prudential and moral significance of affecting people's well-being adversely.

What we primarily want to emphasize here, however, is that NIWA avoids a certain specific problem for CCA, which has recently been identified in the literature (Carlson 2019; 2020; Carlson, Johansson, & Risberg 2021) and concerns the prudential importance of harm. Whatever the exact extent of harm's normative importance, the following principle seems eminently plausible; call it *Prudential Reasons* (PR):

> PR If *a* and *a** are alternative actions open to *S* in a choice situation, and *a* would harm *S* whereas *a** would not, then *S* has some (not necessarily strong or decisive) prudential reason to perform *a** rather than *a*.

As the following case illustrates, however, CCA appears to violate PR:

> *Buttons.* On a board in front of Batman, there are four buttons, B1 to B4, any one of which he can easily press. Pressing B1 would leave Batman very well off; pressing B2 would leave him slightly less well off; pressing B3 would leave him very badly off; and pressing B4 would leave him even worse off. In the nearest possible world where Batman presses B2, it is true that if he had not done so, he would have pressed B1. Further, in the nearest possible world where Batman presses B3, it is true that if he had not done so, he would have pressed B4.[12]

Given CCA, pressing B2 would harm Batman, whereas pressing B3 would not. By PR, then, Batman has some prudential reason to press B3 rather than B2. But it seems clear that he has no such reason— not even a weak *pro tanto* reason.[13] NIWA, on the other hand, escapes this problem. It can hardly be *both* that pressing B2, which would place Batman at a high well-being level, would affect his well-being adversely, *and* that pressing B3, which would place him at a much lower well-being level, would not. On NIWA, then, it is not the case that pressing B2, unlike pressing B3, would harm Batman. Hence, PR is not violated.

Another threat to the normative significance of harm that is also worth considering is posed by the idea that harming is somehow a "contextual" matter.

12. The case is taken from Carlson (2020: 409), with some minor stylistic modifications. Carlson also gives examples of how to make the relevant counterfactuals plausible.

13. It is easy to transform this problem into a *moral* objection to CCA; see Carlson (2019: 797).

This idea is most commonly endorsed by adherents of CCA who believe that counterfactuals are heavily context-dependent (see, e.g., Klocksiem 2012 and Norcross 2005).[14] As Klocksiem puts it, "counterfactuals are highly sensitive to context, and there may be no single way things would have been if things had gone otherwise; [CCA thus] entails that whether something is harmful is as sensitive to context and convention as the counterfactual conditionals that are its backbone" (2012: 289). In other words, according to the contextualist idea, a sentence ascribing harmfulness (or harmlessness) to an event (e.g., 'The Joker's action harms Batman in *Tear Gas*') can be true in one context but false in another, since the proposition expressed differs. This view makes it hard to see how harm could be normatively significant, since sentences such as '*a* is morally wrong' plausibly do not depend on context in this way. After all, an agent can only either do *a* or not do *a*, full stop; she cannot both do *a* relative to one context and not do *a* relative to another context. For this reason, it is hard to see how principles such as PR and various deontological restrictions against harming should even be applied if contextualism about harm is true.

NIWA avoids contextualism about harm: there is no immediate reason to think that whether an event negatively affects someone's well-being depends on context. In light of the plausibility of NIWA, this is good news for the normative significance of harm. It is also good news for NIWA, since there is no independent support for thinking that contextualism about harm is true, and some independent support for thinking that it is false. To see this, let us compare '*e* harms *S*' to a paradigmatically context-sensitive sentence such as '*S* is tall'. If *S*'s height is 190 centimeters, then '*S* is tall' can be true in one context (e.g., when *S* is compared with her colleagues) but false in another (e.g., when *S* is compared with basketball players). Sentences such as '*e* harms *S*' do not behave in this way. For example, even in a context in which deaths and other very severe harms are salient, it can still truly be said of a mosquito bite that it is harmful. Other paradigmatically context-sensitive expressions, such as personal pronouns (e.g., 'I', 'you'), demonstratives (e.g., 'this', 'that'), and temporal indexicals (e.g., 'now', 'today', 'present'), seem to be even worse models for the meaning of 'harm'.

The contextualist threat to the normative significance of harm can thus be dismissed. Accordingly, if CCA implies contextualism about harm (perhaps because the best theory of counterfactuals implies that they are context-sensitive), then this is a further reason to reject CCA.

14. Much the same can be said about so-called *contrastivist* views of harm (see footnote 4). Contrastivism and contextualism are related, since context is usually taken to determine what the relevant contrast event is. Bradley (2012: 408) and Norcross (2005: 171–72) both note that contrastivism and contextualism threaten the moral significance of harm, but neither of them notes (as we do below) that there is independent reason to deny that harm is context-dependent in this way.

3.4. Kinds of Harm

NIWA provides a neat account of two distinctions that are often drawn in the debate about harm: the one between *intrinsic* and *extrinsic* harm, and the one between *pro tanto* and *overall* harm. (Corresponding distinctions can be made with respect to benefit, and what we say below also applies to benefit.)

The distinction between intrinsic harm and extrinsic harm, to begin with, is used to distinguish between events that are harmful *in themselves* and events that are harmful in virtue of their relations to other things. An episode of intense pain is a paradigm example of an intrinsically harmful event; losing one's job is a paradigm example of an extrinsically harmful event. Such examples support the plausible view that an event *e* is extrinsically harmful to *S* just in case it is in some appropriate way related to things that are intrinsically harmful or intrinsically beneficial to *S*. CCA is most commonly thought of as an account of extrinsic harm, and can thus be seen as a suggestion about what the appropriate relation, more exactly, is; namely, a certain kind of counterfactual relation.[15]

Thus understood, however, CCA leaves it mysterious what intrinsic and extrinsic harm *have in common*. It fails to shed light on why both kinds of harm are just that—kinds of one and the same phenomenon: harm. For a counterfactual analysis of intrinsic harm can hardly be correct. For example, it will not do to suggest that an event intrinsically harms a subject just in case it holds, *due to the event's intrinsic features*, that she would have been better off if it had not occurred. That claim has implications to which adherents of CCA surely do not want to commit—for example, that a painful experience, without which the person would have been in even more intense pain, is intrinsically harmless. It is hard to see, then, how a unified account of extrinsic and intrinsic harm could be given if CCA is correct.

If NIWA is correct, however, such a unified account *can* straightforwardly be given. NIWA, as we formulated it above, is in our view a plausible analysis of harm in the general, undifferentiated sense of 'harm' that does not distinguish between intrinsic and extrinsic harm. The distinction can instead be drawn in terms of the way in which the relevant events, more precisely, adversely affect the subject's well-being. When an event, such as an episode of intense pain, is intrinsically harmful to a subject, it not only impacts her well-being negatively but also does so in a very direct, immediate way, which distinguishes it from the impact made by intrinsically neutral but still harmful events such as losing one's job and missing one's flight. Reserving the term 'direct' for this kind of

15. Corresponding things can also be said about CA. However, the details are a bit more complicated with respect to that view, due to the various versions of it that are available, so we shall focus on CCA in the present subsection.

immediate impact that intrinsically bad events have on one's well-being, NIWA motivates the following plausible analysis of intrinsic harm:

> NIWA-I What it is for an event *e* to intrinsically harm an individual *S* is for *e* to directly adversely affect *S*'s well-being.

While direct negative influence on well-being is one kind of negative influence on well-being, it is not the only one. Again, losing one's job and missing one's flight typically negatively influence one's well-being too, though not in the immediate way that is plausibly characteristic of intrinsic harm. Reserving the term 'indirect' for influences that are not direct in this way, NIWA also motivates the following analysis of extrinsic harm:

> NIWA-E What it is for an event *e* to extrinsically harm an individual *S* is for *e* to indirectly adversely affect *S*'s well-being.

Unlike CCA, then, NIWA does provide an answer to the question of what intrinsically harmful and extrinsically harmful events have in common: again, both of them are negative influencers with respect to well-being. The difference simply has to do with what kind of negative influencers they are.

In a similar way, NIWA also provides a neat account both of what *pro tanto* and *overall* harm have in common and of how they differ. This distinction is used to distinguish between events that are harmful *in some respect* or *to some extent*, on the one hand, and events that are harmful, *all things considered*, on the other hand. For instance, a painful operation that significantly improves one's life for many years is a paradigm example of a pro tanto harmful but overall beneficial event.

To account for this distinction, note that we can correspondingly distinguish between a pro tanto and an overall notion of being a negative influence. This distinction has independent support: for example, the Olympic Games might overall influence the economy positively, even if they also influence the economy negatively in some respect (perhaps they cause some people to watch TV instead of going to work). Similarly, an operation that is painful may in that respect be a negative influence on the patient's well-being, even if its later impact on her quality of life means its overall influence was positive. When this holds, NIWA straightforwardly allows one to say that the operation is pro tanto harmful but overall beneficial, which is the intuitively plausible result.

CCA is typically viewed as an account of overall harm: an event's overall harmfulness to someone consists, on this view, in her being *on balance* better off in the nearest world in which the event does not occur. What analysis of pro tanto harm is congenial to CCA? The only answer seems to be this: an event pro

tanto harms a subject just in case she is better off *in some respect* in the nearest world in which the event does not occur.[16] This account of pro tanto harm is implausible, however. In *Tear Gas*, for instance, there is no respect in which Batman would have been better off if the Joker had not performed his action. On the present account, then, the Joker's action does not even pro tanto harm Batman. That cannot be right.

Of course, CCA is technically compatible with other analyses of pro tanto harm. For example, on one such analysis, which avoids the *Tear Gas* problem, a pro tanto harmful event is one that *causes* the person to be in a *state* (in this case, having tear gas in at least one eye) such that she would have been better off in some respect if she had not been in *it*. But any such analysis is going to be *ad hoc*—flowing from something other than CCA itself—and leave it mysterious what pro tanto and overall harm have in common if CCA is true. For example, the alternative analysis of pro tanto harm just mentioned is clearly more congenial to a causal-counterfactual view of harm.

4. Is NIWA Sufficiently Illuminating?

We suspect that despite the virtues of NIWA, some readers might worry that the view somehow fails to accomplish what a successful analysis is meant to accomplish. It is not entirely easy to make this worry more precise, but the general concern is that NIWA should be dismissed as illegitimate because it is in some sense insufficiently illuminating. In this section we shall consider some ways in which this objection can be made more precise, and argue that none of them poses problems for NIWA.

A first possible charge is that NIWA is, in a certain sense, *circular*. Of course, NIWA is not circular in the sense that 'harm' occurs both in the analysans and the analysandum. But, according to this objection, NIWA is nonetheless circular in the sense that saying that *e* negatively affects *S*'s well-being is *just another way* of saying that *e* harms *S*. In other words, according to the objection, that *e* negatively affects *S*'s well-being is just what it *means* that *e* harms *S*.

This kind of objection to a philosophical analysis is not uncommon (which is why we bring it up). But it is very peculiar: if '*e* harms *S*' and '*e* negatively affects *S*'s well-being' really do mean the same thing, then it is hard to see how this could speak *against* NIWA's adequacy, rather than in its favor. (Indeed, if NIWA is to work as a linguistic analysis, '*e* harms *S*' and '*e* negatively affects *S*'s

16. Contrary to what Bradley (2012: 393, n. 8) suggests, appeal to "respects" of well-being does not presuppose pluralism about well-being. For instance, an operation without which the subject would have been on balance hedonically worse off might involve some painful experience that she would have otherwise been spared.

well-being' *must* have the same meaning.) This circularity charge, then, provides no reason to dismiss NIWA.

A slightly more sophisticated complaint is that it is rather 'negatively affects' and 'harms' that mean the same thing, and that NIWA therefore objectionably analyzes harm (to *S*) in terms of harm (to *S*'s well-being). This would be comparable to suggesting that '*S* knows that *p*' means the same thing as '*S* has a true belief that *p* that is caused in the right way', where 'caused in the right way' can in turn only be understood by reference to knowledge. Such a protest is unconvincing for several reasons, however, including the following three. First, 'negatively affects' and 'harms' do not mean the same thing. Even if a poem cannot be harmed, for example, an ill-advised revision can still affect it negatively. Second, surely it is at most in a loose sense of 'harm' that an event might be said to harm someone's *well-being*—a sense that is different from, and arguably parasitic on, the more familiar and strict sense in which an event can harm an *individual* (where the latter is, of course, what NIWA, like CA and CCA, concerns). Saying that an event harms someone's well-being is in this way like saying that her well-being suffers as a result of it. Third, comparing NIWA with the analysis of knowledge as true belief that is caused in the right way is misguided. For unlike the expression 'caused in the right way', the notions that figure in NIWA *can*, as we have urged, be understood independently of the notion that NIWA seeks to analyze. The notion of a negative influence is, again, familiar from elsewhere (see Section 3), and it is surely unproblematic to analyze harm partly in terms of well-being (as all of NIWA, CA, and CCA do). It is at most in a harmless sense, then, that NIWA analyzes harm in terms of harm.

Another, related objection is that NIWA is *uninformative*, or at least insufficiently informative. We would learn nothing of interest by learning that it is true; or at least, everyone already knows that it is true. Indeed, the critic might contend that our elaborate defense of NIWA resembles one of Thomas Nagel's examples of the absurd, in which "someone gives a complicated speech in favor of a motion that has already been passed" (Nagel 1971: 718). NIWA, the critic might suggest, does not itself teach us anything about harm; it is more plausibly seen as something that views like CA and CCA are attempts to elucidate further.

In fact, however, NIWA tells us lots of interesting things about harm, including many things concerning which there is disagreement in the debate. To take just a few examples that we have touched upon already, (i) NIWA vindicates the idea that harm and well-being are closely related; (ii) it shows that harm has something in common with other kinds of negative influences, such as those that have to do with the economy, one's moral worth, and so on; (iii) it vindicates the normative significance of harm; (iv) it reveals that the "attractive first thought" (see Section 1) is not merely a good starting point for theorizing about harm, but straightforwardly provides an analysis of this notion; and not least, (v) it

implies that most leading views of harm, including CCA and CA, are false—and thereby also rules out that those views can successfully elucidate NIWA further (since they are not even extensionally equivalent to NIWA). Rather than NIWA being insufficiently informative, then, several important views of harm have been insufficiently informed by it. Indeed, those who take NIWA to border on the trivial should find it especially problematic for a view to conflict with it. In addition, as we shall argue in the next section, (vi) NIWA also helps rebut Bradley's skepticism about the role of harm in serious theorizing.

Moreover, note that NIWA does not exclude that the notions that it invokes to analyze harm can themselves be analyzed further. While we have not assumed any such analyses, it might be that the notion of well-being, say, can be analyzed in terms of, for example, fitting attitudes, the subject's desires, or the flourishing of human beings. Such an analysis could, compatibly with NIWA, shed further light on what harm is. And the same goes, of course, for the other central notion in NIWA: that of negatively influencing something. For this reason, there is also room in logical space for thinking that this notion can be analyzed in terms of either causal or counterfactual terms, and that NIWA thus might be compatible with the spirit (though not with the letter) of either CA or CCA. While we do not deny that there is such room, we take the discussion in Section 3 to provide strong reason to think that such analyses will nonetheless be very implausible. Thus, we suspect that this room in logical space is little more than just that.

While none of the above objections seems to us to be successful, we do agree that there is nevertheless some sense in which NIWA is comparatively less informative than CA and CCA. Elsewhere, we have speculated that no neat and informative theory of harm might be true (Carlson, Johansson, & Risberg 2021: 173–74), and although NIWA does not vindicate that speculation (since it *is* neat and informative), it might be taken to come quite close. None of this, however, provides any reason to doubt its adequacy. If one had hoped that a more informative analysis of harm would be true, NIWA might be dissatisfying. This might simply mean that we should adjust our expectations.

5. Bradley's Skepticism

Again, some writers in the debate—most notably Ben Bradley (2012)—have expressed skepticism about the suitability of the notion of harm in ethical theorizing. What Bradley claims is not that, contrary to what it seems, we actually do not have prudential and moral reasons to avoid harming ourselves and others (cf. Section 3.3); his view is rather that the concept of harm is so ill-behaved that

we should not state normative claims in terms of it at all.[17] In this section, we shall show how NIWA helps to deal with this skepticism.

According to Bradley, "the notion of harm is a Frankensteinian jumble" (2012: 391). While he does not explain what he means by this, we take his idea to be that harm is a "gruesome" and wildly disjunctive notion.[18] For this reason, Bradley proposes that this notion is also unsuitable in serious ethical theorizing. His main ground for this "messiness thesis" about harm is that all existing non-messy accounts of harm—including, primarily, CCA and various versions of CA—face serious problems. In addition, Bradley also suggests that our judgments about harm are "muddied by moralizing," in that we "are more likely to call an act harmful if we think it is wrong," and proposes that this helps to explain why finding a plausible theory of harm has been difficult (2012: 410).

Independently of NIWA, there is a lot to say about Bradley's reasoning here. One problem is that even if all existing accounts of harm fail, it is hard to see why this would suggest that the notion of harm is gruesome and disjunctive. For one thing, this conclusion does not sit well with the plausible "unity" desideratum that Bradley himself endorses, according to which an analysis of harm "should explain what *all* harms have in common by locating a common core to harm" (2012: 395). Moreover, some of Bradley's objections are that the accounts in question generate *too much* harm. For instance, he notes that CCA implies that Batman's action in *No Golf Clubs* (see Section 3.2) harms Robin and takes this to be problematic, as the more plausible verdict (he thinks) is that Batman's action does not harm Robin. If this is indeed a problem for CCA, it obviously cannot be solved by adding further disjuncts to the counterfactual condition in that theory: that would lead to a view that generates *more*, rather than less, harm.

What about the idea that some of our judgments about harm are "muddied by moralizing" and therefore unreliable? Contrary to what Bradley seems to suggest, this idea, if correct, should come as *good* news for the philosopher who wants to give a (non-messy) analysis of harm. The reason is that this information about the causal influences on our judgments about harm may allow her to "explain away" some of those judgments when they conflict with her view, precisely on the ground that they are moralized and thus unreliable. Hence, if anything, Bradley's conjecture actually speaks against his messiness thesis about harm, rather than in its favor.

17. In this regard, Bradley's skepticism about harm differs from that of, e.g., Holtug (2002). What Holtug seeks to argue is not that the concept HARM is unsuitable as such in moral theorizing, but only that the moral importance of what it stands for is overstated, since no account of harm makes Mill's so-called "harm principle" (according to which, roughly, the state may use coercion only to prevent harm to others) plausible.

18. This interpretation is supported by claims Bradley makes in a different context, to the effect that a kind of pluralism about the concept of excellence implies that that concept is "a gruesome, disjunctive, 'Frankenstein' concept unsuitable for use in moral theorizing" (Bradley 2007).

At any rate, NIWA—which Bradley does not consider—is clearly not a disjunctive account. Moreover, as we have stressed, NIWA also avoids the problems that CCA and CA face. Thus, because a plausible, non-disjunctive account of harm can be given, Bradley's skeptical argument about the role of harm in ethical theory is unsuccessful.

It is worth noting that there may of course be several different *ways* in which an event can negatively influence someone's well-being—and thus, if NIWA is true, several different ways in which an event can harm someone. For instance, as we have argued, some events harm a subject by causally influencing their well-being negatively, whereas other times the relevant relation is grounding rather than causation. More specific and everyday distinctions can also be drawn; for instance, in some contexts we may want to distinguish between physically harming someone (i.e., negatively influencing their well-being by punching them, kicking them, etc.), and psychologically harming someone (i.e., negatively influencing their well-being by making them feel humiliated, threatening them, etc.). However, the fact that there are many different ways of negatively influencing someone's well-being does not suggest that the concept NEGATIVE INFLUENCE ON WELL-BEING is itself disjunctive—that would be like taking the concept MURDER to be disjunctive because there are many ways to murder someone. On the contrary, the concept MURDER is a useful one *precisely because* it captures what many otherwise quite different events—that is, poisoning someone, pushing them out of an airplane, locking them up without water for a sufficiently long time, and so on—have in common; that is, roughly, that they all involve intentionally and wrongly killing someone. Similarly, the fact that events can negatively influence our well-being in many different ways—and thus, on NIWA, harm us in many different ways—is a reason to keep using the concept HARM rather than abandoning it.

A related possible objection to NIWA might be that even though its analysans is at least not explicitly disjunctive, the relevant notion of *negatively affecting*—one of its key elements—fails to be metaphysically "joint-carving" in some appropriate sense; for example, maybe it fails to be metaphysically fundamental, or to play any explanatory role in metaphysics. This could be taken to suggest that NIWA implies, in accordance with Bradley's skepticism, that the notion of harm is a "gruesome" one, which in turn renders it unfit for normative theorizing.

This objection is unconvincing, however, even assuming that the relevant notion of negatively affecting does fail to be metaphysically joint-carving. First, nothing in NIWA rules out that this notion can, in turn, be analyzed in terms of notions that are metaphysically joint-carving. Second, and more important, if the notion of harm is suitable to normative theorizing, this is because harm is *normatively* significant—for instance, because we have prudential reason to avoid harm to ourselves and moral reason not to harm others. There is no reason, independent of this, to think that harm should also be *metaphysically* significant or "joint-carv-

ing." And, crucially, something need not be metaphysically joint-carving in order to be normatively joint-carving, or otherwise normatively significant. After all, if the notion of pain turns out not to be metaphysically fundamental, and to play no explanatory role in metaphysics, this is no threat to its central role in ethics. It is just an implausible ethical claim that pain's normative importance depends on its metaphysical importance—its moral and prudential relevance is due to what it does to the sufferer, rather than what it can do for the metaphysician. And the same, we submit, should be said about harm: its significance to normative theorizing is independent of its significance to metaphysics.

6. Concluding Remarks

In this paper, we have argued that NIWA is the most plausible view of harm, rebutted a class of likely objections to it, and registered some of its important implications.

One question that we have only briefly touched upon concerns the grounds of facts about adversely affecting someone's well-being: when *e* is a negative influence on *S*'s well-being, then in virtue of what does this fact obtain? While a full-fledged treatment of this question is beyond the scope of the paper, we suspect that the most promising answer to it is some form of particularism or pluralism. As we have emphasized, it is plausible that *e* sometimes negatively influences *S*'s well-being at least partly in virtue of certain causal facts (e.g., that *e* causes something intrinsically bad for *S*), other times at least partly in virtue of certain counterfactual facts (e.g., that *S* would have been better off if *e* had not occurred), and yet other times at least partly in virtue of certain other metaphysical facts (e.g., that *e* is a constituent of something intrinsically bad for *S*). While it cannot be ruled out that some neat and unified principle about the grounds of "negative influence" facts that respects this plurality could be given, we also do not see any particularly strong reason to think that this is the case.

This might be yet another disappointing conclusion to certain readers; especially to those who have a preference for tidy philosophical analyses of the form 'to be *F* is to be *G*', where *F* and *G* are on their face quite different kinds of properties. It is thus worth closing on the question of what kind of doxastic attitude such disappointment might motivate. It is true that NIWA leaves many questions unanswered, both about harm and about the notions in terms of which it analyses harm. But it is surely a fallacy to reject an account on the sole ground that it does not immediately provide answers to certain questions—even if those questions are very important ones that concern what the account is about. On the contrary, if the account is otherwise well-supported (as we have argued that NIWA is), the fact that it does not immediately provide answers to those questions might be a

reason to keep investigating them, and to form beliefs about them that are consistent with (and, ideally, supported by) the account, but it is hard to see why it would be a reason to reject the account.

The points just made can be contrasted with an alternative line of thought, according to which a correct analysis of harm will allow us to answer several controversial questions in normative ethics. This idea often seems to be taken for granted in the debate, though it is not always made fully explicit. For instance, some ethical questions that have been approached via the question of what harm is concern the non-identity problem, and other issues in procreative ethics (e.g., Boonin 2014; Gardner 2015; Shiffrin 1999); the moral significance of consent (Shiffrin 1999); Mill's "harm principle" (Holtug 2002); and the plausibility of various deontological doctrines, such as that of doing and allowing harm (e.g., Purves 2019). It would undeniably be nice if an account of harm were to provide answers to these questions, but it is no less undeniable that the truth is not always nice. So perhaps these questions must simply be settled on other grounds. Indeed, as we have stressed, we find it more likely that a correct analysis of harm will rather be one that makes sense of the widespread disagreement about exactly how, why, and to what extent harm is morally important, and thus allows that many competing views about these issues can be viable.

Acknowledgments

For very helpful comments we are grateful to Erik Carlson, Victor Moberger, participants in the Higher Seminar in Practical Philosophy at Lund University, and one area editor and two referees for *Ergo*. Work for this paper was supported by Grants 2018-01361 (Johansson) and 2020-01955 (Risberg) from Vetenskapsrådet.

References

Bontly, Thomas D. (2016). Causes, Contrasts, and the Non-Identity Problem. *Philosophical Studies, 173*(5), 1233–51.

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

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

Bradley, Ben (2007). Review of Robert M. Adams, *A Theory of Virtue*, *Notre Dame Philosophical Reviews*. Retrieved from https://ndpr.nd.edu/news/a-theory-of-virtue-excellence-in-being-for-the-good/

Bradley, Ben (2009). *Well-Being and Death*. Oxford University Press.

Bradley, Ben (2012). Doing Away with Harm. *Philosophy and Phenomenological Research, 85*(2), 390–412.

Carlson, Erik (2019). More Problems for the Counterfactual Comparative Account of Harm and Benefit. *Ethical Theory and Moral Practice, 22*(4), 795–807.
Carlson, Erik (2020). Reply to Klocksiem on the Counterfactual Comparative Account of Harm. *Ethical Theory and Moral Practice, 23*(2), 407–13.
Carlson, Erik and Jens Johansson (2019). Bontly on Harm and the Non-Identity Problem. *Utilitas, 31*(4), 477–87.
Carlson, Erik, Jens Johansson, and Olle Risberg (2021). Well-Being Counterfactualist Accounts of Harm and Benefit. *Australasian Journal of Philosophy, 99*(1), 164–74.
Carlson, Erik, Jens Johansson, and Olle Risberg (2022). Causal Accounts of Harming. *Pacific Philosophical Quarterly, 103*(2), 420–45.
Feit, Neil (2002). The Time of Death's Misfortune. *Noûs, 36*(3), 359–83.
Feit, Neil (2015). Plural Harm. *Philosophy and Phenomenological Research, 90*(2), 361–88.
Feit, Neil (2016). Comparative Harm, Creation and Death. *Utilitas 28*(2), 136–63.
Feit, Neil (2019). Harming by Failing to Benefit. *Ethical Theory and Moral Practice, 22*(4), 809–23.
Gardner, Molly (2015). A Harm-Based Solution to the Non-Identity Problem. *Ergo, 2*(17), 427–44.
Gardner, Molly (2016). Beneficence and Procreation. *Philosophical Studies, 173*(2), 321–36.
Gardner, Molly (2019a). David Boonin on the Non-Identity Problem: Rejecting the Second Premise. *Law, Ethics and Philosophy, 7*, 29–47.
Gardner, Molly (2019b). When Good Things Happen to Harmed People. *Ethical Theory and Moral Practice, 22*(4), 893–908.
Hanna, Nathan (2016). Harm: Omission, Preemption, Freedom. *Philosophy and Phenomenological Research, 93*(2), 251–73.
Hanser, Matthew (2009). Harming and Procreating. In Melinda Roberts and David Wasserman (Eds.), *Harming Future Persons* (179–99). Springer.
Hanser, Matthew (2019). Understanding Harm and Its Moral Significance. *Ethical Theory and Moral Practice, 22*(4), 853–70.
Harman, Elizabeth (2004). Can We Harm and Benefit in Creating? *Philosophical Perspectives, 18*(1), 89–113.
Harman, Elizabeth (2009). Harming as Causing Harm. In Melinda Roberts and David Wasserman (Eds.), *Harming Future Persons* (137–54). Springer.
Holtug, Nils (2002). The Harm Principle. *Ethical Theory and Moral Practice, 5*(4), 357–89.
Johansson, Jens, and Olle Risberg (2018). The Problem of Justified Harm: A Reply to Gardner. *Ethical Theory and Moral Practice, 21*(3), 735–42.
Johansson, Jens and Olle Risberg (2019). The Preemption Problem. *Philosophical Studies, 176*(2), 351–65.
Johansson, Jens and Olle Risberg (2020). Harming and Failing to Benefit: A Reply to Purves. *Philosophical Studies, 177*(6), 1539–48.
Klocksiem, Justin (2012). A Defense of the Counterfactual Comparative Account of Harm. *American Philosophical Quarterly, 49*(4), 285–300.
Nagel, Thomas (1971). The Absurd. *The Journal of Philosophy, 68*(20), 716–27.
Norcross, Alastair (2005). Harming in Context. *Philosophical Studies, 123*(1–2), 149–73.
Northcott, Robert (2015). Harm and Causation. *Utilitas, 27*(2), 147–64.
Parfit, Derek (1984). *Reasons and Persons*. Oxford University Press.
Purves, Duncan (2016). The Harms of Death. In Michael Cholbi (Ed.), *Immortality and the Philosophy of Death* (67–84). Rowman & Littlefield.

Purves, Duncan (2019). Harming as Making Worse Off. *Philosophical Studies, 176*(10), 2629–56.
Rabenberg, Michael (2015). Harm. *Journal of Ethics and Social Philosophy, 8*(3), 1–32.
Shiffrin, Seana (1999). Wrongful Life, Procreative Responsibility, and the Significance of Harm. *Legal Theory, 5*(2), 117–48.
Shiffrin, Seana (2012). Harm and Its Moral Significance. *Legal Theory, 18*(3), 357–98.
Smuts, Aaron (2012). Less Good but Not Bad: In Defense of Epicureanism about Death. *Pacific Philosophical Quarterly, 93*(2), 197–227.
Suits, David B. (2001). Why Death Is Not Bad for the One Who Died. *American Philosophical Quarterly, 38*(1), 69–84.
Thomson, Judith Jarvis (2011). More on The Metaphysics of Harm. *Philosophy and Phenomenological Research, 82*(2), 436–58.
Timmerman, Travis (2019). A Dilemma for Epicureanism. *Philosophical Studies, 176*(1), 241–57.
Velleman, J. David (2008). The Identity Problem. *Philosophy & Public Affairs, 36*(3), 221–44.