This is the author’s accepted manuscript. Please do not cite. The published manuscript can be found at:

<https://www.degruyter.com/document/doi/10.1515/sats-2019-0025/html>

Harm as negative prudential value: A non-comparative account of harm

Abstract

In recent attempts to define ‘harm’, the most promising approach has often been thought to be the counterfactual comparative account of harm. Nevertheless, this account faces serious difficulties. Moreover, it has been argued that ‘harm’ cannot be defined without reference to a substantive theory of well-being, which is itself a fraught issue. This has led to the call for the concept to simply be dropped from the moral lexicon altogether. I reject this call, arguing that the non-comparative approach to defining harm has not been sufficiently explored. I then develop such an account that avoids the difficulties faced by comparative accounts whilst not presupposing a substantive theory of well-being. I conclude that this definition renders a concept of harm that *can* be meaningfully employed in our moral discourse.

Keywords

Harm; defining harm; non-comparative account; negative prudential value; prudential constitution

1. Introduction

The definition of harm has recently come under the spotlight[[1]](#footnote-1), since, despite ubiquitous injunctions against harming in both common-sense morality and ethical theory, surprisingly little has been said about what it is. [[2]](#footnote-2) The meaning of ‘harm’ is generally left up to our intuitions, potentially further exacerbating ethical disputes. A number of definitions have been proposed, but none has gone uncontested. This has led Bradley (2012) to argue that the concept should be dropped from our moral lexicon altogether.[[3]](#footnote-3) I will work from the assumption that Bradley’s call is premature and impracticable. Not only do the shortcomings of current definitions not preclude the possibility of adequately defining harm, but it is hard to know how moral theorizing could proceed without it. Moreover, of the three prominent accounts generally discussed—the non-comparative, temporal comparative, and counterfactual comparative accounts—the latter two have dominated the discussion. I will argue that the merits of such an approach have not been sufficiently explored. This, in addition to the considerable problems faced by comparative approaches, leads me to propose a non-comparative account of harm. I argue that it avoids these problems, accords with our intuitive ascriptions of harm, and does not presuppose a substantive theory of well-being.

1. Groundwork

My focus will be on determining what it is for someone[[4]](#footnote-4) to undergo harm, since the moral significance of ‘harm’ as a concept arguably lies in the fact that we take it to be bad to be harmed. Being harmed is something we want to avoid, and hence we develop prohibitions against it. What it is to *be* harmed is conceptually prior to what it is to *cause* harm. ‘Harming’ here is equivalent to *causing* someone to *be harmed*. For most purposes, this equivalence is sufficient; however, there are instances where this is not the case. The distinction between ‘being harmed’ and ‘harming’ is not always maintained in the literature, which means that it is sometimes not clear whether accounts of harm are accounts of *being harmed* or *causing harm*, nor whether criticisms hinge on their implications for being harmed or harming, further contributing to the conceptual muddle around harm.

Bradley (2012) raises another important distinction—the distinction between ‘intrinsic’ and ‘extrinsic’ harms. Intrinsic harms are harmful for their own sake—e.g. pain. We tend to take pain to be harmful in itself and not because of what it brings about. Extrinsic harms are harmful because of what they bring about. An example that Bradley uses is that of smoking—smoking is harmful because of what it potentially causes, i.e. intrinsic harms. Bradley goes on to claim that intrinsic harms are inherently bad for those experiencing them; hence, claims about intrinsic harms are simply claims about well-being—intrinsic harms entail negative well-being. The problem is that what constitutes well-being is a contentious issue. Bradley thinks that we cannot have a useful concept of ‘intrinsic harm’ that cuts across substantive claims about well-being, making it useless for “axiologically-neutral” moral theorizing (2012, 394). (He sees non-comparative accounts of harm essentially as claims about intrinsic harm.)

Consequently, recent accounts of ‘harm’ are often characterized as attempts to develop an account of extrinsic harm. Here, definitions of harm are proposed that determine what it is to be harmed without reference to the inherent badness of the resultant state. In the resulting comparative theories, a person’s resultant state is compared to a relevant, alternative one. If it can be said to be worse than the alternative, she has been harmed. She is harmed in virtue of being left in a worse state, not in virtue of being left in an intrinsically bad state. Various problems arise. Especially problematic is that comparative theories assume that being left ‘worse off’ is an extrinsic harm in virtue of the fact that this (potentially) brings about some or other intrinsic harm (i.e. ‘being harmed’ essentially consists in having one’s well-being reduced in the appropriate way—temporally or counterfactually). This assumption does not always pan out, and the literature is replete with examples where subjects intuitively seem to have been harmed without being left worse off (temporally or counterfactually) or are left worse off (temporally or counterfactually) without seemingly being harmed.

Surprisingly, there has been little engagement with non-comparative approaches and little, if any, testing of the claim by Bradley (2012) that such an approach cannot avoid presupposing a substantive theory of well-being. Below, I attempt to do just that.

1. The case for a non-comparative account

The two prevalent ‘comparative’ approaches to defining harm are the temporal comparative account and the counterfactual comparative account. The temporal comparative account suggests that a subject (S) is harmed by something if she can be said to be worse off for it than she was prior to it. On the counterfactual comparative account, S’s situation is compared to what it would have been had the ‘something’ not been the case. If S is worse off than she would have been in its absence, she has suffered a harm. Ostensibly, both of these accounts are compatible with different conceptions of well-being—neither definition makes reference to the inherent badness of S’s states. But it is unclear how far this apparent axiological neutrality can take us. Determining whether someone is worse off entails making some kind of evaluation based on what constitutes being better and worse off, which necessitates reference to a substantive theory of well-being. Hence, neither of these comparative accounts deliver a concept of harm that can be consistently applied independently of axiological commitments, calling into question claims that they are axiologically neutral in any useful sense. [[5]](#footnote-5)

Mainly, however, both accounts are criticized for being extensionally inadequate. The temporal comparative account does not allow for harm in cases where someone is already badly off, nor, by extension, for (some) congenital harms (Holtug 2002; Hanser 2008; Thomson 2011; Shiffrin 2012). It also struggles with apparently harmful events that leave their subjects better off (Harman 2004; Shiffrin 1999, 2012),[[6]](#footnote-6) and with “preventative harms”—harms where someone is prevented from receiving a benefit[[7]](#footnote-7) (Holtug 2002; Hanser 2008). Nor can it account for the harm of death since it is unclear how death leaves one ‘worse off’ and some do not consider death a harm (see McMahan (2002)). Shiffrin (1999; 2012) also holds that harming through failing to prevent harm from occurring (“omissions”), cannot be accommodated on this view.[[8]](#footnote-8)

The counterfactual comparative account (in various guises) is often presented as a more viable alternative (e.g. Feinberg 1984; Feldman 1991; McMahan 2002; Thomson 2011; Bradley 2012; Feit 2015; Kloksiem 2013; Petersen 2014; Hanna, 2016; Purves 2018). Here, S is harmed if she finds herself worse off for an event than she would have been had it not occurred. This allows for harm to those who are already badly off, those who are left better off (in some instances), and those whose situation remains unchanged.[[9]](#footnote-9) It is also claimed that counterfactual comparison allows for preventative harm, and harm through omission and death.[[10]](#footnote-10) (Nevertheless, work still needs to be done to determine which counterfactual existence particular deaths are to be compared to.[[11]](#footnote-11) ) Even if all of these claims are conceded, there still remain seemingly legitimate instances of harm that the view cannot accommodate. Congenital harming remains problematic in light of Parfit’s origin view[[12]](#footnote-12) (Hanser 2008; Shiffrin 2012). Parfit’s (1987) non-identity problem also arises, where the very harming event allows for the existence of the harmed person in the first place.

Two further problems are especially intractable for the counterfactual comparative account: 1) It entails that harm occurs in ‘failure-to-benefit’ cases (Shiffrin 1999, 2012; Bradley 2012; Petersen 2014; Purves 2018), and 2) it is unable to deal with instances of overdetermination and pre-emption (McMahan 2002; Norcross 2005; Hanser 2008; Bradley 2012; Shiffrin 2012; Petersen 2014; Johansson 2017). In failure-to-benefit cases, someone is harmed every time that he is not benefitted when he could have been (e.g. any stranger I come across is harmed when I do not give him the spare $10 in my pocket). The pre-emption problem arises when two intuitively harmful events befall a subject closely together, where the second has either a similar or worse outcome for her than the first. The first event is rendered not-harmful due to the existence of the second. [[13]](#footnote-13);[[14]](#footnote-14)

In light of the problems faced by comparative accounts of harm, one would think that non-comparative approaches have been comprehensively refuted, especially since these seem quite intuitive—someone is harmed because of what happens to her, regardless of her foregoing or counterfactual situation. Surprisingly, there is very little by way of consciously-developed, non-comparative definitions of harm. Moreover, such approaches are generally more or less dismissed out-of-hand.

Two accounts are usually mentioned when there is talk of non-comparative accounts: those of Harman (2004, 2009) and Shiffrin (1999, 2012) (e.g. Hanser 2008; Bradley 2012; Gardner 2015, 2017[[15]](#footnote-15)). Essentially, both consider S to be harmed when she is caused to fare badly. Faring badly is ‘bad in itself’, and, accordingly, the focus is on whether S is badly off irrespective of her prior or counterfactual states. The most basic non-comparative accounts simply list examples of harms rather than defining the concept (e.g. Harman (2004) with “pain, early death bodily damage and deformation”). This leads to the criticism that such accounts cannot unify ‘harm’, as they do not specify what all of the items on the list have in common (e.g. Bradley, 2012). Harman (2009) tries to unify her list by stipulating that S is harmed if he is caused to be in a bad state. Inevitably, this elicits counterexamples where someone is caused to be in a bad state but is not plausibly harmed (e.g. Hanser 2008; Gardner 2015). Furthermore, Harman’s formulation cannot distinguish between pro tanto and overall harm, where S is caused to be in a bad state but ends up being better off overall (Bradley, 2012).[[16]](#footnote-16) The bad state is a pro tanto harm, but not an overall one. Harman also seems unable to allow for preventative harms and the harm of death.

The other prominent non-comparative account discussed is that of Shiffrin (1999, 2012), who defines harm as that which runs counter to what S rationally wills. Here, harm “involves conditions that generate a significant chasm or conflict between one’s will and one’s experience, one’s life more broadly understood, or one’s circumstances” (1999, 123). Bradley (2012) rejects this approach on the grounds that 1) one can rationally will to be harmed, and 2) Shiffrin’s definition entails that only rational creatures can be harmed.[[17]](#footnote-17)

It should be clear that the main stumbling blocks for non-comparative approaches are: 1) providing a unified conception of harm that includes all entities liable to harm, 2) distinguishing between pro tanto and overall harm, and 3) accounting for preventative harms and death. Yet, as discussed above, the currently-favored counterfactual comparative account struggles with congenital harms, the non-identity problem, the failure-to-benefit problem, and the pre-emption problem. Work also needs to be done to accommodate the harm of death. Moreover, it is not clear that this account is robustly axiologically neutral in any useful sense. Hence, it makes sense to explore the non-comparative approach more comprehensively.

1. Prudential value

One of the problems faced by non-comparative accounts is providing a unified description of harms that also allows for the possibility that something may be a *pro tanto* but not an overall harm. We can overcome this problem by changing the perspective from which harms are attributed, which will also allow us to unify harms. Arguably, the most prominent feature that all instances of being harmed have in common is that it is *bad* in some way for *those affected by it*. Whatever else we mean when we say that S has been harmed, we also mean that something bad has happened *to S*. Hence, ‘bad’ here should be understood in a subject-relative sense. The ‘badness’ of bad things here lie in what they hold for their subjects. Not all bad things are bad in this way. One may describe a movie as ‘bad’ but this is not to say that it is bad *for* anyone. This much is relatively uncontroversial—harm consists in such things that are bad for those liable to it. However, we still need to account for why ostensibly bad things may not be bad for S in a particular context. We can do this by stipulating that harms are also ‘bad for’ S in a *subjective* *sense*: a harm is something that is bad for S *from S’s own perspective*. Thus, harms are bad because those subjected to them take them to be bad.[[18]](#footnote-18) This claim is not as implausible as it may at first appear. Consider death. Even though death is often considered to be a great harm, it may not come as a harm for someone who is suffering from a terminal illness and who wants to die. We often allow for the possibility that some things come as harms to some but not to others. This is why consent plays such an important role when it comes to assessing the harmfulness of some events. We allow for the possibility that someone may consider the negative effects of something bad (e.g. pain) to be outweighed by its benefits, which makes it not-bad *for them*. A surgeon does not harm a patient that has consented to surgery in light of its benefits. If the surgeon were to operate against the patient’s will, she would be harming him, since she is inflicting something that is bad *for him*.[[19]](#footnote-19) This is what lies behind the sense that we need to be able to distinguish between pro tanto versus overall harms in the first place. Some bad states may not bad *for us*, because they facilitate something that we value more. Thus, by shifting our focus to *subjective badness* we are able to both unify harms and to allow for pro tanto versus overall harm. As a first pass, then, my claim is being harmed entails being subjected to that which we subjectively disvalue.

At this stage, my account may seem similar to that of Shiffrin. However, my claim is *not* that only beings who can rationally will can be harmed. As mentioned in footnote 19 above, my claim is that all entities that can be described as ‘value-fixing kinds’ can be harmed. For a being to ‘value’ a state of affairs, it needs to have welfare—it must be possible for things to go better or worse for it. A cat has welfare in this way; a pebble does not. However, such a being need not ‘rationally will’ its own faring in the sense of cognitively assessing a situation and explicitly entertaining a desired state of affairs in any anthropomorphic sense. We usually take beings to have welfare when they exhibit behavior that indicates preferences or aversions. Mice try not to be eaten. It does not seem a stretch to say that they ‘disvalue’ being eaten. We can say that being eaten is *bad* *for* them, without thereby presupposing that they somehow entertain explicit, rational preferences in this regard. Mice simply need to tend to avoid being eaten—and perhaps show distress when they do face such a situation—for this claim to be plausible. To avoid the unnecessary (and illegitimate) cognitive connotations when talking about valuing and disvaluing in this way, my claim can be phrased as follows: being eaten holds *negative prudential value* for mice.

References to that which holds negative prudential value for value-fixing-kinds should not be taken to refer only that which is *inherently* bad for them. I am not claiming that harm entails being subjected to situations with given intrinsic “bad-making-properties”[[20]](#footnote-20) (although this may often be the case with non-human value-fixing-kinds). Something can hold negative prudential value for S irrespective of its inherent good-making and bad-making properties. Something at t can have particular intrinsic good-making properties for S and still come as a harm to S. For example, S may take pleasure to be good-making. Yet, this does not mean something pleasurable cannot come as a harm to S. S may be presented with a delicious meal, which provides her with great gastronomical pleasure. Yet, unbeknownst to S, her meal is non-vegan, causing S to be subjected to something that holds negative prudential value for her—eating animal products. S’s gastronomical experience is both intrinsically good and a prudentially bad for her.[[21]](#footnote-21) Hence, I distinguish between intrinsic ‘goods’ and ‘bads’ and prudential ‘goods’ and ‘bads’. Whereas intrinsic goods and bads are those things that we value or disvalue for their own sake, prudential goods or bads constitute a broader category—these are things that hold value for us for a variety of reasons, both intrinsic and extrinsic. The value that prudential goods and bads hold for us does not necessarily correlate with how they affect our well-being.[[22]](#footnote-22)

The objection may be raised that I’m merely presenting a desire-satisfaction view of well-being. This is not the case. As already explained, with prudential bads I am not limiting harms to that which is inherently bad for us, and I am leaving room for the possibility that that which is inherently bad for us may not ultimately be prudentially bad. For us My claim is thus not that to be harmed is necessarily to experience negative well-being. Moreover, I am not claiming that to have our desires met is to experience positive well-being. If anything, I am presenting a preference-based view of what it is to be harmed, if we were to use a very broad conception of ‘preference’, which encompasses all of the things that we value or might value without necessarily (currently or ever) being aware of our ‘preferences’.[[23]](#footnote-23)

Provisionally, I propose the following account of being harmed:

HARMED: To be subjected to something that holds negative prudential value for one.

HARMED remains neutral about the reasons for particular disvaluations as well as whether or not such disvaluations are justified. This implies that S may end up being mistaken as to whether or not a given state of affairs is good or bad for him, from an objective perspective.[[24]](#footnote-24) Such states of affairs remain *subjectively,* *prudentially* good or bad for S, at least until he is given occasion to think otherwise.[[25]](#footnote-25) The corollary is that someone who is fully informed about a particular state of affairs cannot be mistaken regarding the harm that it affords him, unless there is reason to believe that he does not have the capacity to make such a judgement. Hence, I disagree with Bradley’s claim that one can rationally will to be harmed.

The question remains as to whether HARMED is extensionally adequate. But before discussing this, I want to assess the strengths of HARMED.

1. Strengths

HARMED avoids the problems faced by both other non-comparative accounts and by comparative accounts. It provides a unified account of harm without entailing that only rational beings can be harmed. A value-fixing-kind need not have a cognitive understanding of a situation or a rational reason for taking it to hold negative prudential value for it. If S can experience pain or distress and if S tends to avoid these, pain and distress can plausibly be said to hold negative prudential value for S. This implies that a range of animals can be liable to harm.[[26]](#footnote-26) Yet, this also allows for the fact that humans are liable to a wider range of harms than other value-fixing-kinds, since humans are able to conceive of and value a wider range of goods and bads. This seems right, as we readily allow that humans can be harmed in ways that animals cannot (e.g. being offended). In addition to unifying harms, HARMED also allows us to distinguish between pro tanto versus overall harm. An intrinsic bad need not be harmful overall to S if this is outweighed by its prudential goodness. Rescuee is not harmed by Rescuer who breaks his arm in order to rescue him because the prudential goodness of having his life saved outweighs the intrinsic badness of the pain.

HARMED is able to deal with both death and preventative harms. Death comes as a harm to those for whom it holds negative prudential value—whether due to instinct or existential reasons. It also allows for the possibility that death does *not* come as a harm to some and may come as a harm even to those who are very badly off. Some may prudentially value death, e.g. for bringing an end to suffering. On the other hand, it also explains why death is often thought to be a very grave harm—arguably, value-fixing-kinds tend to experience a strong, biological drive to survive (or, perhaps more accurately, for their genes to survive). *Ceteris paribus*, they ‘want’ to go on existing and resist anything that causes them not to.

Similarly with preventative harms. S is harmed when her life-saving medication is stolen because this causes a state of affairs that holds negative prudential value for her—not having her life saved. One may want to counter that the true disvaluable state here is that S dies. Yet, nothing prevents us from delineating various harms and degrees of harm. The primary harm here may be that S dies. But it seems wrong to say that being prevented from having her life saved is inconsequential for S. S is also harmed in as far as being prevented from having her life saved is bad for her. Hence, the thieves harm S in as far as they cause her to suffer the derivative harm of not having her life saved but not in as far as she suffers the harm of dying. This seems right. While they cannot be said to be entirely responsible for the harm that befalls S, they also cannot be said to have played no hand in it.

HARMED also fares well when it comes to the other problems faced by the comparative accounts. Congenital harms are harming in as far the resultant circumstances hold negative prudential value for those affected, regardless of their comparative states.[[27]](#footnote-27) In addition, pre-emptive and overdetermined harms are not necessarily rendered harmless.[[28]](#footnote-28)

Two of our problem cases remain: omissions and failures-to-benefit. Difficulties here come down the fact that both involve failures to act rather than overt actions. It is not obvious that Bystander harms Pedestrian in failing to save him from being hit by a bus, since much here turns on how we conceive of ‘cause’. There are two possible responses, whilst retaining a largely intuitive notion of ‘cause’: i) concede that Bystander does not harm Pedestrian, since the disvalued state is caused by the bus and not Bystander; or, ii) specify the disvalued state(s) that results from Bystander’s actions more carefully, and hold him responsible for harming Pedestrian accordingly. Bystander may thus be responsible for *failing to prevent harm* and thus harms Pedestrian to the extent that this inaction holds negative prudential value for Pedestrian.[[29]](#footnote-29) Neither option is immediately counterintuitive, which reflects the ambivalence that often arises with omissions. Morally speaking, it seems wrong to not prevent harm from occurring when one can. Nevertheless, a plausible definition of harm must be *amoral* and thus not affected by our intuitions regarding wrongness (see Bradley, 2012).[[30]](#footnote-30)

It is not obvious that the failure-to-benefit problem even arises on HARMED. On the counterfactual comparative account, taken as an account of ‘being harmed’, most of us are harmed most of the time, since others constantly fail to benefit us in some way. Taken as an account of ‘harming’, most of us harm others most of time, since there are a myriad ways in which we fail to benefit others at any given time. The implausibility here stems from the fact that we do not take harm to entail states of affairs that routinely arise from our everyday dealings with one another and which we usually do not take to be bad. On HARMED, failures-to-benefit are harms only in as far as not being thus benefitted holds negative prudential value for those affected. I would contend that most of the ways in which others can benefit us do not entail the kinds of states of affairs the lack of which routinely holds negative prudential value for us. It might be nice if the diner at the next table buys me a meal, but not being bought a meal by my fellow diners does not usually hold negative prudential value for me.

There are a subset of failure-to-benefit cases where it seems plausible to claim that harm occurs. These are cases where S is already badly off and where Z fails to help whilst being able to. If S is on the brink of starvation Z fails to buy him a meal, Z may be said to have harmed S, as starving to death holds negative prudential value for S. Yet, this is an instance of failing to prevent harm, rather than a failure-to-benefit, making it similar to Pedestrian, above. Z is not responsible for the harm that befalls S to the extent that he would have been had he been responsible for starving S. As with Bus, the harmfulness of his inaction increases to the extent that S had the expectation that Z would save him.

There may be instances where S, rather eccentrically, *does* prudentially disvalue not being bought a meal by random fellow diners. This will entail both that S is harmed and that her fellow diners harm her. I will say more on the issue of seemingly trivial harms below. For now, note that trivial or eccentric harms are only problematic if we assume that harming is always wrong—i.e. if we already hold a moralized conception of harm. But, as already discussed, our definition should be *amoral*. Hence, it is not problematic that our definition allows for harms that are not obviously morally wrong. The wrongness of particular harms will be established in terms of particular moral theories. Moreover, any given moral theory needs to distinguish between justified and unjustified harms (more on this below). Hence, we may conclude that the harm in this instance is justified, and that S is not wronged.

HARMED manages to avoid the problems faced by the other accounts. But there are obvious objections that can be levelled against it: It may be extensionally inadequate in that it seemingly allows for trivial harms and it does not allow for ‘false negatives’ (harms that subjects do not disvalue). There is also the difficulty of determining what holds prudential disvalue for subjects with limited cognitive capacities and/or who are unable to communicate.

1. Objections

HARMEDentails that: 1) one cannot be subjected to a state of affairs that holds negative prudential value for one without being harmed, and 2) one cannot suffer a harm if one is not subjected to an event or state of affairs that holds negative prudential value for one. Both of these are potentially problematic.

Entailment 1) allows for the criticism that the definition overextends the concept of harm in that someone may prudentially disvalue something trivial from a third-party perspective. As discussed, this is only problematic if we already moralize the concept of harm. On an amoral understanding, the possibility of picking out seemingly trivial harms becomes less problematic. Moreover, this possibility can be a strength of HARMED, as it allows for the great variety of ways in which value-fixing-kinds, especially humans, can be harmed. Think of an agoraphobic who experiences distress when being made to step outside. To most, stepping outside may seem trifling. Yet, it is plausible to say that it comes as a harm Agoraphobic. The harm may be justified (perhaps this is a radical form of therapy); however, this does not detract from Agoraphobic’s experience of harm. Moreover, 1) need not undermine HARMED’s value as a normative concept—all that is needed to preserve the credibility of a normative theory that moralizes harm is to develop a concept of *justified harm*, which most moral theories already do.[[31]](#footnote-31) This allows subjective harm claims to be morally evaluated in terms of whether or not the harm inflicted is justified. The details of how this is to be fleshed out fall under the scope of particular moral theories.

Nevertheless, it is true that HARMED does not allow for value-fixing-kinds themselves (again, primarily humans) to distinguish between *subjectively* trivial disvalued states and harms. Hence, it needs to be amended so as to include only that which is (subjectively) ‘significantly’ disvalued. This does not preclude the possibility things which are apparently trivial can count as harms, but it does allow for a subjective distinction between mere inconveniences and irritations, and harm.

Entailment 2) is more problematic: S cannot suffer a harm if the event or state of affairs at issue does not hold significant prudential disvalue for S. It may be objected that this under-ascribes harm, as it seems possible to harm beings in ways that they cannot appreciate. This could be due to epistemological, cognitive, or constitutive limitations, as one may find in the cases of children, the severely cognitively impaired, animals, and perhaps even extreme cases of can be called ‘false consciousness’ (where someone is cognitively able to make credible harm-judgements but has severely limited access to epistemological facts or is under psychological duress).[[32]](#footnote-32) It is also possible to subject rational adults to things that they have simply never encountered (or that have no evolutionary precedent). (Do I prudentially disvalue having my browsing history run through a particular algorithm?) These objections need not be problematic for HARMED, partly because it does not place a time limit on harm judgements. In cases of harm to children, false consciousness, and as-yet-unencountered harms, harm judgements can be made retroactively, given the development of the cognitive capacity to value such events and/or access to relevant information. Hence, it is possible to suffer harm while not immediately appreciating it. But this is not yet entirely satisfactory. It is not plausible to say that someone who is subjected to sexual abuse as a child is not harmed until such time as she is able to appreciate the nature and ramifications of the situation. Moreover, there are cases such as those of the severely cognitively impaired, animals (to varying degrees), and ‘intractable’ false-consciousness cases where the situation may never be appreciated. This problem can partially addressed through the provision made above for non-cognitive ‘valuing’. A value-fixing-kind (especially animals) can instinctively disvalue something without needing to have a cognitive grasp of it.[[33]](#footnote-33)

Yet, we are still faced with the problem of apparent harms that those affected are unable to appreciate, given their particular situation or the nature of the harm. Someone who is severely cognitively impaired, for example, can be harmed in ways that he cannot appreciate. He may be exploited for medical research, even while not experiencing any discomfort. Whereas the subjectivity of HARMED is a strength in avoiding the problem of objectively developing a conception of ‘bad for’ that covers all possible instances of harm across all value-fixing-kinds, it now faces the problem of instances where someone is unable to appreciate when things are plausibly going badly for him. Looking at the nature of value-fixing-kinds can be of help here. As already argued, one can experience harm if one is a value-fixing-kind—if one has welfare—which means that one has the kind of constitution that allows things to go better or worse for one. For all value-fixing-kinds, this constitution is due to evolutionary heritage and (non-cognitive) ontogenetic factors. Some value-fixing-kinds (humans) are also prudentially constituted by cognitive ontogenetic factors—they come to value various things for culturally-acquired reasons. Call all of these ‘reasons’ that cause value-fixing-kinds to prudentially value and disvalue given things their *prudential constitution*. This will allow for expanding and refining HARMED in order to better reflect the ways in which things can hold negative prudential value for all value-fixing-kinds.

Given the discussion above, HARMED can be refined as follows:

HARMED: Value-fixing-kind S is harmed by an event or state of affairs (E) iff

1. S knows about E and E holds significant negative prudential value for S, or
2. S will find out about E in the future and E will then hold significant negative prudential value for S, or
3. If S *were* to know about E and/or if S *were* able to make the relevant kinds of value judgments, E would hold significant negative prudential value for S, given S’s prudential constitution.

HARMED now entails that situations that S does not know about or does not fully appreciate can still hold (significant) negative prudential value for S, either because S will disvalue the situation (when more information becomes available to S) or S will counterfactually disvalue the situation (in the closest possible world where S is able to appreciate the situation, given S’s prudential constitution). The kinds of harms that can befall S are constrained by the kind of being that S is (phylogenetically and ontogenetically), and the closest possible world is where S is able to fully appreciate the situation without being a different kind of being (or a different person). This formulation is admittedly vague, and in cases where S cannot communicate, we essentially need to make an educated guess as to S’s prudential constitution, given what we know about S.[[34]](#footnote-34) Nevertheless, it is an improvement on the current situation, where we attribute harm based on unanalysed intuitions and/or particular theories of well-being. On HARMED, our guess may be wrong, but we will have some grounds for claiming that it is justified, given what we know about S or beings like S.

A problem that remains arises from not having a time-limit on harm judgements: What happens if specific harm judgements by fully-informed, cognitively-competent adults are later retracted? S may judge an event as harmful at t1 but change his mind at t2. This need not be problematic. There is no reason, in principle, why someone’s harm attributions need to stay constant. S may come to new insights by t2 and conclude he was wrong at t1. Retroactively revised harm judgements can become problematic when someone has been held accountable for the harm at t1 and where S’s reassessment at t2 holds significant moral (and perhaps legal) implications for him.

It is important to clarify just what such a retraction would look like. These would not be instances where someone has learned to live with the consequences of a harm. Take Harman’s (2004) case of a rape victim (S) who comes to love the child that results from the rape. Does her loving her child at t2 entail that she was not harmed at t1? It need not. It is possible for something to come as a harm to someone even if it ends up rendering her benefits. S can disvalue the rape and resultant trauma even while loving her child and believing her life to be better because of the child.[[35]](#footnote-35) HARMED allows for a fine-grained distinction where S can both significantly disvalue the rape and value the child. The rapist has still harmed S. Yet, could S possibly reassess the events at t1 and decide, given what she knows at t2, they were not a harm? In principle, she can, which would render events at t1 not-harmful. S may decide that this particular child would not have been born had the rape not occurred, and hence that, on balance, she was not harmed.[[36]](#footnote-36) This seems incongruous, but the incongruity may be due to our intuitions regarding the likelihood that such a reassessment will be made, given the grievousness of the harm here, rather than the possibility of the reassessment itself. There are cases where reassessment appears less incongruous. Say my employer requires me to have an annual influenza vaccine, which I object to on the grounds that this violates my right to autonomy. The annual shots come as a harm to me. One year, an especially deadly strain of influenza breaks out, causing many deaths among the unvaccinated. Upon surviving, I reassess my judgement and conclude that the mandatory vaccinations were not a harm after all. This reevaluation is less counterintuitive. Sometimes the way that events unfold casts new light on earlier experiences. It seems right that we are able reevaluate our judgements accordingly.

1. Conclusion

It appears a plausible non-comparative definition of harm is possible. HARMED is not only able to unify harms, but can also accommodate those problem cases that the alternative accounts struggle with. It allows for the intuition that being harmed somehow entails having things go badly for one, without necessitating that only that which is inherently bad is harmful. It can distinguish between pro tanto and overall harm and does not presuppose a substantive theory of well-being. It entails that various kinds of entities can be harmed, but limits the type of harm that can befall them relative to the kinds of entities they are. Finally, HARMED also explains why harms are the kinds of things that we tend to moralize—they are *bad for* those affected by them and are better avoided. If HARMED is adequate, it allows for the retention of ‘harm’ as a moral concept. Given the ubiquity of the concept in our moral discourse and its intuitive importance, this is preferable to dropping the concept altogether.

References:

Bradley, B. (2004). When Is Death Bad for the One Who Dies? *Nous, 38*(1), 1-28.

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

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

Feinberg, J. (1984). *The Moral Limits of the Criminal Law: Harm to Others* (Vol. One). Oxford: Oxford University Press.

Feinberg, J. (1992) Wrongful Life and the Counterfactual Element in Harming, in *Freedom and Fulfilment: Philosophical Essays*. Princeton: Princeton University Press, 3–37.

Feit, N. (2015). Plural Harm. *Philosophy and Phenomenological Research, XC*(2). doi:10.1111/phpr.12033

Feit, N. (2017). Harming by Failing to Benefit. *Ethical Theory and Moral Practice*. doi:10.1007/s10677-017-9838-6

Feldman, F. (1991). Some Puzzles About the Evil of Death. *The Philosophical Review, 100*(2), 205-227. Retrieved from http://www.jstor.org/stable/2185300

Gardner, M (2015) A harm based solution to the non-identity problem. *Ergo*, *An Open Access Journal of Philosophy*. 2: 17. pp. 427-444. DOI: 10.3998/ergo.12405314.0002.017

Gardner, M (2017) When Good Things Happen to Harmed People. Ethical Theory and Moral Practice. https://doi.org/10.1007/s10677-017-9840-z

Gert, B. (1998). *Morality: Its Nature and Justification.* Oxford: Oxford University Press.

Hanna, N. (2016). Harm: Omission, Preemption, Freedom. *Philosophy and Phenomenological Research, 93*, 251-273. doi:doi:10.1111/phpr.12244

Hanser, M. (2008, September). The Metaphysics of Harm. *Philosophy and Phenomenological Research, LXXVII*(2).

Harman, E. (2004). Can We Harm and Benefit in Creating? *Philosophical Perspectives*, 89-113. Retrieved from http://www.jstor.org.ez.sun.ac.za/stable/3840929

Harman, E. (2009). Harming as causing harm. In M. A. Wasserman (Ed.), *Harming Future Persons* (pp. 137--154). Springer Verlag.

Holtug, N. (2002). The Harm Principle. *Ethical Theory and Moral Practice, 5*(4), pp. 357-389.

Johansson, J. and Risberg, O. (2019) The Preemption Problem. *Philosophical Studies* 176(2), 351-365. https://doi.org/10.1007/s11098-017-1019-x

Kloksiem, J. (2012) A defence of the counterfactual comparative account of harm. *American Philosophical Quarterly, 49*(4), 285-300. Retrieved from http://www.jstor.org/stable/23460806

McMahan, J. (2002). *The Ethics of Killing. Problems at the Margins of Life.* Oxford: Oxford University Press.

Norcross, A. (2005). Harming in Context. *Philosophical Studies: An International Journal for Philosophy in the Analytic, 123*(1/2), 149-173. Retrieved from http://www.jstor.org/stable/4321577

Parfit, D. (1987). *Reasons and Persons.* Oxford: Oxford University Press.

Petersen, T. S. (2014). Being Worse Off: But in Comparison with What? On the Baseline Problem of Harm and the Harm Principle. *Res Publica , 20*(2), 199-214.

Purshouse, C. (2016), A Defence of the Counterfactual Account of Harm. *Bioethics*, 30: 251-259. doi:[10.1111/bioe.12207](https://doi.org/10.1111/bioe.12207)

Purves, D. (2016). Accounting for the Harm of Death. *Pacific Philosophical Quarterly, 97*(1), 89-112. doi:DOI: 10.1111/papq.12031

Purves, D. (2018) Harming as Making Worse Off. *Philosophical Studies*, https://doi.org/10.1007/s11098-018-1144-1

Shiffrin, S. V. (1999). Wrongful life, procreative responsibility, and the significance of harm. *Legal Theory, 5*, 117–148. doi:10.1017/S1352325299052015

Shiffrin, S. V. (2012). Harm and Its Moral Significance. *Legal Theory, 18*, 357–398. doi:10.1017/S1352325212000080

Thomson, J. J. (2008). *Normativity.* Chicago: Open Court.

Thomson, J. J. (2011). More On The Metaphysics of Harm. *Philosophy and Phenomenological Research, 82*(2), 436-458. doi:http://www.jstor.org/stable/23035326

1. E.g. Feinberg 1984; Parfit 1987; Shiffrin 1999, 2012; Harman 2004, 2009; Norcross 2005; Hanser 2008, 2013; Bradley 2012; Thomson 2011; Kloksiem 2012; Feit 2015, 2017; Gardner 2015; Hanna 2016; Purshouse 2016; Purves 2018. [↑](#footnote-ref-1)
2. See Gert (1998) on the priority that common morality gives to harm and Shiffrin (2012) for a discussion of the role that harm plays in ethical theory. [↑](#footnote-ref-2)
3. Norcross (2005) makes a similar point. [↑](#footnote-ref-3)
4. Anthropocentric terms should be taken as referring to all entities liable to harm. [↑](#footnote-ref-4)
5. Also see Norcross (2005). [↑](#footnote-ref-5)
6. Harman’s (2004) is the classic case here where someone undergoes lifesaving, but invasive surgery. [↑](#footnote-ref-6)
7. E.g., life-saving medication is stolen en route to hospital, unbeknownst to the dying recipient. [↑](#footnote-ref-7)
8. Even if Bystander were deliberately to fail saving Pedestrian from being hit by a bus, he does not harm him, since it is the bus that leaves Pedestrian worse off than before. [↑](#footnote-ref-8)
9. Although see Norcross (2005). [↑](#footnote-ref-9)
10. Feinberg (1984); Feldman (1991); McMahan (2002); Bradley (2012); Feit (2015) [↑](#footnote-ref-10)
11. See, e.g., Feldman (1991), McMahan (2002), Bradley (2004), and Purves (2016). [↑](#footnote-ref-11)
12. Here one’s identity necessarily relates to one’s genetic make-up such that it holds across all possible worlds, and hence one could not have been born without a genetic congenital harm without being a different person. [↑](#footnote-ref-12)
13. Norcross (2005) gives the example of Bobby Knight who crushes Philosopher’s windpipe rather than dismembering her, which he would otherwise have done. [↑](#footnote-ref-13)
14. In overdetermination cases, two such events occur simultaneously and both are rendered not-harmful due to the existence of the other. [↑](#footnote-ref-14)
15. Gardner (2015) develops her own “existence account” of harm which is essentially a hybrid between the non-comparative and counterfactual accounts. She attempts to allow for divergent substantive theories of well-being with a clause that specifies that condition T is a harm to S if there is a way in which S can be better or worse off to T *in some respect*. Bradley (2012) argues that talking about respects of well-being in this way already presupposes a pluralistic theory of well-being and thus violates axiological-neutrality, and, accordingly, I will disregard theories that make such a move (also see Hanser, 2008). [↑](#footnote-ref-15)
16. The classic example that of Feinberg (1992) where a rescuer breaks a trapped rescuee’s arm in order to save him from drowning. [↑](#footnote-ref-16)
17. As I will discuss below, I disagree with 1). [↑](#footnote-ref-17)
18. This implies that only entities who *can* value states of affairs can be harmed. Contrast this with Thompson (2011) who claims that all “goodness-fixing-kinds” (see Thomson, 2008) can be harmed. Arguably, Thomson’s category of “goodness-fixing-kinds” is much too wide, in that it implies that we can harm not only human beings and cats but also toasters, since there is a “good way” to be a toaster. On my view, we cannot harm a toaster, but we *can* harm human beings and other animals. Call such beings: “value-fixing-kinds”. To my mind, we lose nothing in saying that toasters can be damaged but not harmed. On the other hand, with this approach, we gain the insight that harm has to do with that which is valuable to us. [↑](#footnote-ref-18)
19. While consent makes someone’s valuing of something clear, we also allow for cases where consent may be unobtainable. In an emergency, the surgeon may be permitted to operate without consent, since we assume that the patient would have consented, given the value that people generally attach to their lives. [↑](#footnote-ref-19)
20. See Bradley (2009) where he discusses well-being in terms of the value of one’s life being determined by the particular “good-making” and “bad-making” properties that obtain in it. [↑](#footnote-ref-20)
21. S’s reasons for disvaluing the eating of animal products may ultimately come down to that which S intrinsically disvalues. The thought of the suffering of the animals involved in food production may cause S psychological distress. Nevertheless, this does not undermine the extrinsic nature of the harm that S suffers here, nor does it preclude the possibility that S disvalues the practice for reasons other than her own well-being. [↑](#footnote-ref-21)
22. I suspect that if we take value-fixing-kinds to be arranged in a continuum, from cognitively least sophisticated to cognitively most sophisticated, it will mostly be intrinsic bads that hold negative prudential value for those on the lower end of the scale. Humans, with greater cognitive capacity to entertain alternative possibilities and future outcomes, will have more extrinsic bads. [↑](#footnote-ref-22)
23. ‘Preference’ here would be synonymous with ‘desire’, ‘inclination’, ‘tendency toward’, ‘instinct’ and all of the ways in which value-fixing-kinds manifest partiality and predisposition. More on this below. [↑](#footnote-ref-23)
24. Smoking may be bad for S from an objective perspective, but S may persist in according it positive prudential value. As such, it is not a harm for S. [↑](#footnote-ref-24)
25. S may find that he was misinformed about the health effects of smoking and thus revise the prudential value he accords it when becoming aware of this. On the other hand, he may not care. [↑](#footnote-ref-25)
26. As will be discussed, the source of their ‘valuation’ does not matter. A state of affairs can hold negative prudential value for a being purely due to its evolutionary heritage and ontogenetic make up. [↑](#footnote-ref-26)
27. This raises the question of how to determine whether or not future people will be harmed by our actions. In essence, we would need to guess at the value future people will accord their circumstances, given what we know about beings like them. [↑](#footnote-ref-27)
28. In *Bobby Knight*, Philosopher is harmed in as far as having her windpipe crushed holds negative prudential value for her, irrespective of what Knight might have done. [↑](#footnote-ref-28)
29. This opens the possibility that, were Bystander known to Pedestrian, the harm could be greater in proportion to how much his betrayal of Pedestrian’s trust holds negative prudential value for Pedestrian. [↑](#footnote-ref-29)
30. Of course, it is open to proponents of the temporal comparative account to make the same argument against this criticism. [↑](#footnote-ref-30)
31. As with self-defense and the punishment of criminals, for example. [↑](#footnote-ref-31)
32. Consider a remote island community that has accepted sexual activity with young children for generations and where such children either never get to know anything different or are entirely beholden to their abusers. Even as adults, their harm judgements may be considered problematic. [↑](#footnote-ref-32)
33. From a third-party perspective, prudential value “cues” such as avoidance and distress can give us reason to believe that such a being is experiencing harm. [↑](#footnote-ref-33)
34. Say S is a slug. We can infer that S can be harmed, at minimum, by being caused physical damage, by being caused to go hungry, and by being killed, given what we know about animals in general and slugs in particular. Animals tend to do their utmost to avoid these things. It seems unlikely that slugs can be subject to offence; hence, this does not form part of S’s prudential constitution. Nevertheless, we can revise this assumption in future, if warranted. [↑](#footnote-ref-34)
35. This case is different from the surgery above in that the rape is not a pro tanto harm, but an overall one. [↑](#footnote-ref-35)
36. The harm at t1 becomes pro tanto. [↑](#footnote-ref-36)