



Tanya de Villiers-Botha*

Harm as Negative Prudential Value: A Non-Comparative Account of Harm

<https://doi.org/10.1515/sats.2019-0025>

Published online September 22, 2020

Abstract: In recent attempts to define “harm,” comparative accounts of harm, specifically counterfactual comparative accounts, have been touted as the most promising approaches to defining the concept. Nevertheless, such accounts face serious difficulties. This has led to the call for the concept to simply be dropped from the moral lexicon altogether. I reject this call, arguing that *non-comparative* approaches to defining harm have not been sufficiently explored. I develop such an account and claim that it avoids the difficulties faced by comparative accounts while not presupposing a substantive theory of well-being, which is taken as a key failing of non-comparative accounts. 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 since, despite ubiquitous injunctions against harming in both common-sense morality and ethical theory, surprisingly little has been said about what it is.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

¹ For example, (Bradley 2012; Feinberg 1984; Feit 2015, 2019; Gardner 2015; Hanna 2016; Hanser 2008; Harman 2004, 2009; Kloksiem 2012; Norcross 2005; Parfit 1987; Purshouse 2016; Purves 2019; Shiffrin 1999, 2012; Thomson 2011).

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.

*Corresponding author: Tanya de Villiers-Botha, Department of Philosophy, University of Stellenbosch, Private Bag X1, Matieland, Stellenbosch 7602, South Africa,
E-mail: tdev@sun.ac.za. <https://orcid.org/0000-8790-9062>

uncontested. This has led Bradley (2012) to argue that the concept should be dropped from our moral lexicon altogether.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 – non-comparative, temporal comparative, and counterfactual comparative accounts – the latter two have dominated the discussion. I contend that the merits of non-comparative approaches have not been sufficiently explored. I propose a non-comparative account of harm that avoids the problems faced by comparative accounts and accords with Bradley's (2012) desiderata for a satisfactory definition of harm.

2 Groundwork

My focus will be on determining what it is for someone⁴ to be subjected to harm, since the moral significance of “harm” as a concept arguably lies in the fact that we take it to be bad to be in a harmed state. 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.

Bradley (2012) raises another important distinction – that 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. Bradley uses the example 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, for him, claims about intrinsic harms are simply claims about well-being – intrinsic harms entail negative well-being. The problem in terms of defining harm is that there are different, rival theories as to what constitutes well-being. If we want a concept of harm that is not limited to a specific theory of well-being, we cannot make any substantive claims about what constitutes having one's well-being negatively affected. For Bradley (2012), such “axiological neutrality” is one of the main criteria for a satisfactory definition of harm that will be of use in our moral theorizing. Hence, the only options that we seem to be left with are general formulations to the effect that being harmed constitutes having one's well-being negatively affected (whatever that may entail), or having something intrinsically bad happen to one, or being made to fare badly. Problematically, such formulations

³ Norcross (2005) makes a similar point.

⁴ Anthropocentric terms here refer to all entities liable to harm.

stipulate *sufficient* conditions for harm to have occurred, but not *necessary* ones, potentially making such formulations extensionally inadequate. Such accounts also do not allow for another important distinction that needs to be taken into account when defining harm: that between *pro tanto* harms and “all-things-considered” harms, meaning that they count as harms events that leave us badly off in the short term, but which are beneficial overall (life-saving surgery is the classic example of *pro tanto* harm). Hence, when defining harm in terms of intrinsic badness *simpliciter*, it is easy to come up with counterexamples of being made to fare badly that do not plausibly constitute harm (See Section 3 below for a more detailed discussion.).

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 such *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, then 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.⁵ Various problems also arise here. Especially problematic is that such comparative theories assume that being left “worse off” is an extrinsic harm, since this 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).⁶ This assumption does not always pan out, and the literature is replete with examples where subjects 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.

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 referring to intrinsic badness, necessitating an appeal to a substantive theory of well-being. Below, I attempt to rectify this situation.

3 The Case for a Non-Comparative Account

The two prevalent *comparative* approaches to defining harm are temporal comparative accounts (in various guises) and counterfactual comparative

⁵ For example, S is harmed by being hit by a bus, since, *ceteris paribus*, such an event leaves her *worse off* in terms of just about any plausible dimension one might want to take into account when considering how an event affects a person. S is *not* harmed by virtue of being left in an intrinsically bad state.

⁶ The difference between temporal and counterfactual accounts will be discussed in greater detail in Section 3. For current purposes, what is germane is that both focus on whether or not someone is left worse off for an event, as is illustrated in the bus example above.

accounts (also in various guises). Temporal comparative accounts suggest 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 counterfactual comparative accounts, S's situation is compared to what it would have been had the "something" not obtained, irrespective of her prior state. Hence, if S is *worse off* for an event than she would have been in its absence, she has suffered a harm. Thus, as mentioned, comparative accounts focus on whether someone is left *worse off* for an event (in some appropriate way) rather than the badness of the state in which the event leaves the person. Ostensibly, both of these approaches are compatible with different conceptions of well-being – neither makes reference to the inherent badness of S's resultant 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 as to what constitutes being better and worse off, which at some point necessitates reference to a substantive theory of well-being. And there is no guarantee that different theories of well-being will converge in terms of when someone is considered to be worse off. 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.⁷

Nevertheless, both approaches are mainly criticized for being extensionally inadequate. Temporal comparative accounts do not allow for harm in cases where someone is already badly off, nor, by extension, for (some) congenital harms (Hanser 2008; Holtug 2002; Shiffrin 2012; Thomson 2011).⁸ They also struggle with apparently harmful events that leave their subjects better off (Harman 2004; Parfit 1987; Shiffrin 1999, 2012),⁹ and with "preventative harms" – harms where someone is prevented from receiving a benefit¹⁰ (Holtug 2002; Hanser 2008). Nor can they account for the harm of death, since it is unclear how death leaves one "worse off" (as opposed to in no state at all), and some do not consider death a harm at all (see McMahan (2002)). Shiffrin (1999; 2012) also argues that harming through failing to prevent harm ("omissions"), cannot be accommodated on this view.¹¹

Consequently, counterfactual comparative accounts are often presented as more viable (e.g., Bradley 2012; Feinberg 1984; Feit 2015; Feldman 1991;

⁷ Also see Norcross (2005).

⁸ These are cases where S is born into a state that would be considered harmful if it were to befall her later in life (e.g., blindness). Since S is not worse off than before – S is born in this state – S is not harmed.

⁹ Parfit's (1987) is the classic case here where someone undergoes lifesaving, but invasive surgery.

¹⁰ For example, life-saving medication is stolen en route to hospital, unbeknownst to the dying recipient.

¹¹ For example, 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.

Hanna 2016; Kloksiem 2012; McMahan 2002; Petersen 2014; Purves 2019; Thomson 2011). 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 for an event.¹² It is also claimed that counterfactual comparison accommodates preventative harm and harm through omission and death.¹³ (Nevertheless, work still needs to be done to determine which counterfactual existence particular deaths are to be compared to.¹⁴) Yet, even if all of these claims are conceded, there still remain legitimate instances of harm that the view cannot accommodate. Congenital harm remains problematic in light of Parfit's *Origin View*¹⁵ (Hanser 2008; Shiffrin 2012). Related to this is Parfit's (1987) non-identity problem, where the very harming event allows for the existence of the harmed person in the first place.¹⁶

Two further problems are intractable for the counterfactual comparative account: (1) It entails that harm occurs in "failure-to-benefit" cases (Bradley 2012; Petersen 2014; Purves 2019; Shiffrin 1999, 2012), and (2) it is unable to deal with instances of overdetermination and pre-emption (Bradley 2012; Hanser 2008; Johansson and Risberg 2019; McMahan 2002; Norcross 2005; Petersen 2014; Shiffrin 2012). In failure-to-benefit cases, someone is harmed whenever he is not benefitted when he could have been.¹⁷ The pre-emption problem arises when two intuitively harmful events befall a subject closely together the first event is rendered not-harmful due to the existence of the second.¹⁸ In overdetermination

12 Although see Norcross (2005).

13 Feinberg (1984), Feldman (1991), McMahan (2002), Bradley (2012), and Feit (2015).

14 See, e.g., Feldman (1991), McMahan (2002), Bradley (2004), and Purves (2016).

15 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 (Parfit 1987).

16 Here, the very harming event is responsible for the person's existence, the person has a life worth living, and given the Origin View, the person could not have existed in the absence of the harming event. For example, a couple conceive a child while one of them is ill, knowing full well that the child's chances of being born with a disability are thereby increased. The child is born with a disability, yet her life is worth living. If they had not conceived at that time, *this* child would not have been born. Hence, *she* is not worse off than she would otherwise have been and is not harmed.

17 For example, any stranger I come across is harmed when I do not give him the spare \$10 in my pocket.

18 For example, Norcross's (2005) example of Bobby Knight who is angry at Philosopher and is intent on dismembering him. Due to successful anger-management techniques, Knight ends up crushing Philosopher's windpipe instead. On counterfactual comparative accounts, Philosopher is not worse off than he would have been for his crushed windpipe since he would otherwise have been dismembered.

cases, two ostensibly harmful events occur simultaneously, and both are rendered not-harmful due to the existence of the other.¹⁹

In light of the problems faced by comparative accounts of harm, one would think that non-comparative approaches have been comprehensively refuted, especially since they seem 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 addressing non-comparative accounts: Harman's (2004, 2009), and Shiffrin's (1999, 2012) (e.g., Bradley 2012; Gardner 2015, 2019; Hanser 2008²⁰). Essentially, both consider S to be harmed when she is caused to fare badly. Faring badly is "bad in itself" irrespective of S's prior or counterfactual states. 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 these items have in common (e.g., Bradley 2012). Harman (2009) tries to unify her list by simply stipulating that S is harmed if she is caused to be in a bad state. This elicits counterexamples where someone is caused to be in a bad state but is not plausibly harmed.²¹ Furthermore, Harman's formulation cannot distinguish between *pro tanto* and overall harm (Bradley 2012).²² The bad state is a *pro tanto* harm, but not an overall one. Harman also seems unable to allow for preventative harms (it is not the preventative event that causes the person affected to be badly off) and the harm of death (one seemingly cannot be badly off when one is dead).

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.

¹⁹ For example, Batman has a fatal heart attack. He is also hit by a cannon ball that is certain to kill him. Neither event leaves him worse off than he would otherwise have been (adapted from Bradley 2012).

²⁰ Gardner (2015) develops her own "existence account" which is a hybrid between 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 already presupposes a pluralistic theory of well-being and thus violates axiological neutrality. Accordingly, I will disregard theories that make such a move.

²¹ Hanser (2008) gives the example of Nobel Prize Winner, who loses cognitive function as result of a stroke. Although she is worse off than that she was when she had full cognitive function, she still has average cognitive function. We usually do not consider having average cognitive function to be bad – most people find themselves in this position. Hence, she is not harmed by her stroke.

²² A classic example is Feinberg's (1992) where a rescuer breaks a trapped rescuee's arm in order to save him from drowning.

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” Shiffrin (1999, 123). Bradley (2012) rejects this approach, since (1) one can rationally will to be harmed and (2) Shiffrin’s definition entails that only rational creatures can be harmed.²³

Clearly, 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, 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 is also needed 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.

4 Prudential Value

One of the problems faced by non-comparative accounts is providing a unified description of harm that also allows for *pro tanto* harm. We can overcome both problems by changing the perspective from which harms are attributed. Arguably, the most prominent feature that all instances of harm have in common is that they are *bad for those affected*. 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 lie in what they hold for their subjects. Not all bad things are bad in this way. A movie can be “bad” but not bad *for* anyone. This much is 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 are sometimes *not* bad for S in a particular context (i.e., *pro tanto*). 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.²⁴ This

²³ As I will discuss below, I disagree with (1).

²⁴ This implies that only entities who *can* value states of affairs can be harmed. Contrast this with Thomson (2011) who claims that all “goodness-fixing-kinds” (see Thomson, 2008) can be harmed. Thomson’s category of “goodness-fixing-kinds” is much too wide; 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.” We lose nothing by saying that toasters can be damaged but not harmed. On the other hand, we gain the insight that harm has to do with that which is valuable to us.

claim is not as implausible as it may seem. 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. That is why consent plays such an important role when 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 wishes, she would be harming him, since she is inflicting something that is bad *for him*. This is what lies behind the sense that we need to be able to distinguish between *pro tanto* and overall harms in the first place. Some bad states may not be bad *for us*, because they facilitate something that we value more. 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 that 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 24 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 entertaining a desired state of affairs. 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”²⁵ (although this may often be the case with non-human

²⁵ 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.

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. 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 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.²⁶ Thus, 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 – 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.²⁷

It may be objected that this is simply 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. 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, using 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.”²⁸

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 on the reasons for particular disvaluations as well as on whether or not such disvaluations are justified. This implies that S may end up

²⁶ 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 may cause S psychological distress. This does not undermine the extrinsic nature of the harm that S suffers here, nor preclude the possibility that S disvalues the practice for reasons other than her own well-being.

²⁷ 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.

²⁸ “Preference” here would be synonymous with “desire,” “inclination,” “tendency toward,” “instinct,” and all the ways in which value-fixing-kinds manifest partiality and predisposition.

being mistaken as to whether or not a given state of affairs is good or bad for her from an objective perspective.²⁹ Such states of affairs remain *subjectively, prudentially* good or bad for S, at least until she is given reason to think otherwise.³⁰ The corollary is that someone who is fully informed about a state of affairs cannot be mistaken regarding the harm that it affords him, unless he does not have the capacity to make such a judgment. 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 first, I will assess the strengths of HARMED.

5 Strengths

HARMED avoids the problems faced by both other non-comparative accounts and comparative accounts. As opposed to the non-comparative accounts discussed, it provides a unified account of harm without entailing that only rational being 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 S tends to avoid these, pain and distress can plausibly be said to hold negative prudential value for S. This implies that a wide range of animals are liable to harm.³¹ 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 bads. This seems right; we readily allow that humans can be harmed in ways that animals cannot (e.g., being offended). HARMED also allows us to distinguish between *pro tanto* and overall harm. An intrinsic bad need not be overall harmful to S if its harm is outweighed by its prudential goodness. Rescuer is not harmed by Rescued who breaks his arm in order to rescue him, because the prudential goodness of having his life saved outweighs the badness of the pain.

HARMED is able to deal with both death and preventative harms (as opposed to both non-comparative accounts and temporal comparative accounts). Death comes as a harm to those for whom it holds negative prudential value – whether due to instinct or existential reasons, such as the fear of death. It also allows for the possibility that death does *not* come as a harm to some and may come as a harm

²⁹ Objectively, smoking may be bad for S, but S may persist in according it positive prudential value. As such, it is not a harm for S.

³⁰ S may find that she was misinformed about the health effects of smoking and revise the prudential value she accords it.

³¹ 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.

even to those who are very badly off. Some may prudentially value death, e.g., for bringing an end to suffering. Yet, 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. *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 actual 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. 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.

HARMED also fares well when it comes to other problems faced by comparative accounts. Congenital harms are harmful in as far the resultant circumstances hold negative prudential value for those affected, regardless of their comparative states (i.e., S may disvalue being blind, even if she is not worse off for it). In addition, contra counterfactual comparative accounts, pre-emptive and over-determined harms are accommodated.³²

Two problem cases remain: omissions and failures-to-benefit. Difficulties here come down the fact that both involve failures to act rather than actions. It is not obvious that Bystander harms Pedestrian in failing to save him from the bus, since much turns on how we conceive of “cause.” There are two possible responses while retaining an intuitive notion of “cause”: (i) concede that Bystander does not harm Pedestrian, since the disvalued state is caused by the bus, 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 be responsible for *failing to prevent harm* and thus harms Pedestrian to the extent that this inaction holds negative prudential value for Pedestrian.³³ 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

³² In *Bobby Knight* (see footnote 18), Philosopher is harmed in as far as having her windpipe crushed holds negative prudential value for her, irrespective of what Knight might have done. Batman (footnote 19) is harmed in as far as he disvalues dying, irrespective of a second fatal event.

³³ 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.

plausible definition of harm must be *amoral* and thus not affected by our intuitions regarding wrongness (see Bradley 2012).³⁴

It is not obvious that the failure-to-benefit problem even arises on HARMED. On counterfactual comparative accounts, taken as accounts of “being harmed,” most of us are harmed most of the time, since others constantly fail to benefit us in some way. Taken as accounts of “harming,” most of us harm others most of the time because 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. On HARMED, failures-to-benefit are harms only in as far as not being thus benefitted holds negative prudential value for those affected. I 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 subsets 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 despite being able to. If S is on the brink of starvation and Z fails to buy her a meal, Z may be said to harm S, in as far 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 her.

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 about 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).

HARMED manages to avoid problems faced by the other accounts. But there are obvious objections against it: It may be extensionally inadequate in that it

³⁴ Of course, it is open to proponents of the temporal comparative account to make the same argument against this criticism.

seemingly allows for trivial harms, and it does not allow for “false negatives” (apparent 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. I turn to these objections next.

6 Objections

HARMED can be objected to on the following grounds: Entailment (1), one cannot be subjected to a state of affairs that holds negative prudential value for one without being harmed, allowing for trivial harms, and Entailment (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, rendering obvious harms non-harmful. I will address these objections in detail below.

Entailment (1) allows for the criticism that HARMED overextends the concept of harm in that someone may prudentially disvalue something seemingly trivial from a third-party perspective. As mentioned, 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 variety of ways in which value-fixing-kinds can be harmed. Consider 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 to Agoraphobic. The harm may be justified (perhaps this is a form of therapy); however, this does not detract from Agoraphobic’s *experience* of being harmed. Furthermore, Entailment (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 allow for.³⁵ This allows subjective harm claims to be morally evaluated in terms of whether or not the harm is *justified*.

Nevertheless, it is true that HARMED does not allow for value-fixing-kinds themselves (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 harms.

³⁵ As with self-defense and the punishment of criminals, for example.

Entailment (2) is more problematic: S cannot suffer a harm if a given event or state of affairs 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 what can be called “false consciousness” (where someone is cognitively able to make credible harm-judgments but has severely limited access to epistemological facts or is under psychological duress).³⁶ 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, partly because HARMED does not place a time limit on harm judgments. In cases of harm to children, false consciousness, and as-yet-unencountered harms, harm judgments 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 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. There are also cases such as the severely cognitively impaired and animals (to varying degrees) where the situation may never be appreciated. This problem can partially be 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.

Yet, we are still faced with the problem of apparent harms that those affected are unable to appreciate, given their 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 developing an objective 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, value-fixing-kinds, have the kind of constitution that allows things to go better or worse for them. For all value-fixing-kinds, this constitution is due to evolutionary heritage and (non-cognitive) ontogenetic

³⁶ 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 judgments may be considered problematic.

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
- a. S knows about E and E holds significant negative prudential value for S, or
 - b. S will find out about E in the future and E will then hold significant negative prudential value for S, or
 - c. 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 negative prudential value for S, either because S will disvalue the situation (when more information becomes available) or because 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.³⁷ 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 judgments: What happens if specific harm judgments by fully informed, cognitively competent adults are later retracted? S may judge an event as harmful at t1 but change her 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 she was wrong at t1. Retroactively revised harm

³⁷ Say S is a slug. We can infer that S can be harmed, at a 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 offense; hence, this does not form part of S’s prudential constitution.

judgments *may* become problematic when someone has been held accountable for the harm at t_1 and where S 's reassessment at t_2 holds significant moral (and perhaps legal) implications for him.

It is important to clarify just what such a retraction might look like. 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 t_2 entail that she was not harmed at t_1 ? 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. 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 t_1 and decide, given what she knows at t_2 , they were not a harm? In principle, she can, which would render events at t_1 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. Nevertheless, this change of heart does not nullify the harm experienced by S between t_1 and t_2 . Furthermore, it need not have legal implications – S 's rapist still committed a crime and is being punished accordingly, and subjective harm attributions are not the only factors that contribute to the illegality of actions. A case such as this may seem 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, 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, a deadly strain of influenza breaks out, causing many deaths among the unvaccinated. Upon surviving, I reassess my judgment 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 to reevaluate our judgments accordingly.

7 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 relativizes the type of harm that can befall them 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?”. *Noûs* 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. 1, Oxford: Oxford University Press.
- Feinberg, J. 1992. “Wrongful Life and the Counterfactual Element in Harming.” In *Freedom and Fulfilment: Philosophical Essays*, 3–37. Princeton: Princeton University Press, <https://doi.org/10.1017/S0265052500000467>.
- Feit, N. 2015. “Plural Harm.” *Philosophy and Phenomenological Research* 90 (2): 361–88.
- Feit, N. 2019. “Harming by Failing to Benefit.” *Ethical Theory & Moral Practice* 22: 809–23.
- Feldman, F. 1991. “Some Puzzles About the Evil of Death.” *Philosophical Review* 100 (2): 205–27.
- Gardner, M. 2015. “A harm Based Solution to the Non-Identity Problem.” *Ergo, An Open Access Journal of Philosophy* 2 (17): 427–44.
- Gardner, M. 2019. “When Good Things Happen to Harmed People.” *Ethical Theory and Moral Practice* 22: 893–908.
- Gert, B. 1998. *Morality: Its Nature and Justification*. Oxford: Oxford University Press.
- Hanna, N. 2016. “Harm: Omission, Preemption, Freedom.” *Philosophy and Phenomenological Research* 93 (2): 251–73.
- Hanser, M. 2008. “The Metaphysics of Harm.” *Philosophy and Phenomenological Research* 77 (2): 421–50.
- Hanser, M. 2013. “Harm.” In *The International Encyclopedia of Ethics*, edited by H. LaFollette, 2299–307. Hoboken: Blackwell Publishing Ltd.
- Harman, E. 2009. “Harming as Causing Harm.” In *Harming Future Persons*, edited by M. A. Wasserman, 137–54. Springer. https://doi.org/10.1007/978-1-4020-5697-0_7.
- Harman, E. 2004. “Can We Harm and Benefit in Creating?”. *Philosophical Perspectives* 18 (1): 89–113.
- Holtug, N. 2002. “The Harm Principle.” *Ethical Theory and Moral Practice* 5 (4): 357–89.
- Johansson, J., and O. Risberg. 2019. “The Preemption Problem.” *Philosophical Studies* 176 (2): 351–65.
- Kloksiem, J. 2012. “A Defence of the Counterfactual Comparative Account of Harm.” *American Philosophical Quarterly* 49 (4): 285–300.

- 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 Tradition* 123 (1/2): 149–73.
- 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 (4): 251–9.
- Purves, D. 2016. "Accounting for the Harm of Death." *Pacific Philosophical Quarterly* 97 (1): 89–112.
- Purves, D. 2019. "Harming as Making Worse Off." *Philosophical Studies* 176: 2629–56.
- Shiffrin, S. V. 1999. "Wrongful Life, Procreative Responsibility, and the Significance of Harm." *Legal Theory* 5: 117–48.
- Thomson, J. J. 2008. *Normativity*. Chicago, Open Court.
- Shiffrin, S. V. 2012. "Harm and Its Moral Significance." *Legal Theory* 18: 357–98.
- Thomson, J. J. 2011. "More on the Metaphysics of Harm." *Philosophy and Phenomenological Research* 82 (2): 436–58.