

ARTICLE

# The Harm Principle and the Nature of Harm

Anna Folland 

Uppsala University, Uppsala, Sweden  
Email: [anna.folland@filosofi.uu.se](mailto:anna.folland@filosofi.uu.se)

## Abstract

This article defends the Harm Principle, commonly attributed to John Stuart Mill, against recent criticism. Some philosophers think that this principle should be rejected, because of severe difficulties with finding an account of *harm* to plug into it. I examine the criticism and find it unforceful. Finally, I identify a faulty assumption behind this type of criticism, namely that the Harm Principle is plausible only if there is a full-blown, and problem-free, account of harm, which proponents of the principle can refer to.

## 1. Introduction

Arguments for restricting the state from exercising power over individuals in the society often appeal to the so-called Harm Principle.<sup>1</sup> According to this principle, commonly attributed to John Stuart Mill, the state can justifiably intervene against an individual only if intervening prevents harm. Philosophers appeal to the Harm Principle to argue that the state cannot justifiably intervene against, for example, in vitro fertilization (IVF), surrogate motherhood, expressions of sexuality or religious beliefs, and publication of controversial books (Holtug, 2002). Since these acts or activities are harmless, some argue, the state cannot legitimately intervene against them.

Philosophers have widely debated the soundness of the Harm Principle (see for instance Bell, 2020; Brink, 2013; Brown, 2010; Browne, 2016; Dripps, 1998; Edwards, 2014). In this article, I focus on a kind of criticism that has recently emerged. According to this criticism, we should question the plausibility of the Harm Principle, due to difficulties with finding an account of *harm* that can be plugged into it (Holtug, 2002; Jonas, 2016; Petersen, 2014). There are various accounts of the nature of harm in the literature (for an overview and a critical discussion, see Bradley, 2012). As it turns out, some argue, there is no conception of harm such that the Harm Principle, when that conception is plugged into it, amounts to a plausible and useful principle. If the critics are right, then the arguments that appeal to the Harm Principle to protect individual liberty all fail. As of yet, however, there is no systematic discussion of this type of criticism. In this article, I defend the Harm Principle against this criticism.

<sup>1</sup>The Harm Principle is also – some think preferably – called the Liberty Principle (Westmoreland, 2020).

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.

The article proceeds as follows. I begin by providing some preliminaries, in [section 2](#). In [section 3](#), I present the criticism of the Harm Principle, which, as I argue in [section 4](#), is unenforceable. I argue that even the strongest form of the criticism, which surprisingly has been neglected by the critics, fails to refute the Harm Principle. In [section 5](#), I identify a mistaken assumption behind this type of criticism, namely the assumption that the Harm Principle is plausible only if its proponents can refer to a full-blown, and problem-free, account of the nature of harm. [Section 6](#) concludes.

## 2. Preliminaries

It might be helpful to begin with a precise formulation of the Harm Principle. This is easier said than done. Much has been said both about how Mill should be interpreted and about how contemporary proponents should formulate the principle. With that said, I suggest the following formulation.

HP: If state intervention against an individual is justified, then this is at least partly because it prevents harm.<sup>2</sup>

The principle states a *necessary*, but not a sufficient, condition for when state intervention is justified. Harm prevention is not necessarily the entire reason why state intervention is justified on HP, but it is at least part of the reason. Other formulations of the principle often leave out the explanatory element of HP – that is, they leave out the “because” feature. Although such formulations do express that harm prevention is necessary for justified state intervention, they do not express that harm prevention can *explain* why the state is so justified. But those inclined to accept the Harm Principle are likely to have explanatory intentions too.

The debate about Mill’s principle is vast. My objective is not to examine the criticism against HP in general. Rather, I focus on whether certain issues regarding the nature of harm pose any special problems for the plausibility of HP. Therefore, I limit the discussion in this article in various ways.

Philosophers debate whether the best version of HP allows *paternalistic* intervention – that is, intervention that prevents an individual from harming *themselves* (Brink, 2013, ch. 55; Dworkin, 1972; Persson, 1998; Saunders, 2016). However, I do not discuss cases of (potential) self-harm in this article. Neither do I discuss whether the principle should include only *direct* harm, or indirect harm too. Piers Norris Turner illustrates the difference: “If my drinking ... leads me to perform some further action that directly offends others, the drinking itself ... only indirectly offended others” (2014, 319). I intend to exclusively discuss cases that intuitively count as direct harms. Furthermore, I focus on

---

<sup>2</sup>In *On Liberty*, Mill famously presents the so-called Harm Principle as “one very simple principle”. Despite its alleged simplicity, his presentation of the principle remains a topic of interpretation: “The object of this essay is to assert one very simple principle, as entitled to govern absolutely the dealings of society with the individual in the way of compulsion and control, whether the means used be physical force in the form of legal penalties or the moral coercion of public opinion. That principle is that the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number is self-protection. That the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinions of others, to do so would be wise or even right” (Mill [1859] 1977, 223–24).

harm to actual human persons, but do not take a stand on whether harms to animals or merely possible beings also count. I also remain neutral on whether only *non-consensual* harm should count, which would mean that the principle exclusively refers to harm that the victim herself has not agreed to (Brink, 2013; Saunders, 2016).

I bracket the discussion about whether the principle should be read *ex ante* (where the risk of harm is relevant) or *ex post* (where actual harm is relevant) (Holtug, 2002, 359–60). I will not discuss any cases in which merely the risk of harm is at stake. Unless otherwise explicitly stated, I remain neutral regarding whether we should favor an *origin-centered* reading of the principle, which requires that the individual from whom the harm originates is identical to the individual against whom the state can justifiably intervene. On the alternative, *origin-neutral*, reading the state may justifiably coerce an individual whose actions do not relate to the harm (see Holtug, 2002, for a discussion of merits and demerits of origin-neutral and origin-centered versions).

Finally, and as my formulation of HP suggests, I exclusively discuss an *outcome*-based version of the principle, as opposed to *justification*-based versions. What matters on the former is whether intervening against an individual prevents a harmful outcome, and not whether the decision-makers' motivation, or purpose, for intervening is to prevent harm (Holtug, 2002, 362–63).

### 3. Difficulties with finding a suitable conception of harm

According to a recent stream of criticism, no conception of harm is such that HP, when that conception is plugged into it, amounts to a plausible and useful principle (Holtug, 2002; Jonas, 2016; Petersen, 2014). The critics consider various theories of harm in combination with HP, and argue that even the best candidate fails. If they are correct, HP risks being either implausible or useless.

A prominent idea is that the nature of harm is comparative, meaning that harmful events make individuals *worse off*. As there are different ways of spelling out the comparative baseline, there are different comparative accounts of harm. Holtug (2002), Petersen (2014) and Jonas (2016) take different comparative accounts into consideration, and argue that none yields a plausible version of HP when plugged into it.

One comparative account, which appeals to a temporal baseline, is the Temporal Comparative Account (TCA):

TCA: An event *e* harms a subject *s* if, and only if, *e* makes *s* worse off (in terms of welfare) after *e* than *s* was prior to *e*.

According to TCA, harms are events that make subjects worse off than they were prior to the event. Now, let *HP + TCA* denote the conjunction of HP and TCA. Petersen and Holtug argue that *HP + TCA* is implausible, because of a problem with so-called *preventive harms*. Such events are harmful in virtue of preventing someone from being in a better state. Holtug provides an example: "Suppose that a person's pain would have gone away, had I not acted to ensure that it continues. Clearly, I harm him, despite the fact that I leave him in no more pain than he was prior to my intervention" (2002, 368). *HP + TCA* counterintuitively rules out justified state intervention in cases of this kind, since the act ensures that someone's pain continues without making him worse off than he was before.

According to Jonas (2016), HP + TCA also mistakenly rules out justified state intervention in situations where someone is born with a bad condition.<sup>3</sup> Jonas (2016, 393) illustrates the problem with the case of Amy, who is 3 years old and suffers from severe asthma from infancy. Amy's condition is caused by exposure to Environmental Tobacco Smoke (ETS), as Amy's mother smoked throughout pregnancy and her two parents both smoke in her presence since birth. We should assume, I suppose, that before Amy got her condition she had a lower wellbeing level or perhaps *no* wellbeing level. (One might think that Amy did not exist before she got the condition, and that only existing individuals have a wellbeing level.) If so, HP + TCA counterintuitively rules out that the state can justifiably intervene in this case, since Amy was not better off before she got her condition.

Holtug (2002) and Peterson (2014) both consider, and eventually reject, the idea that HP might successfully be combined with a comparative account of harm that employs "the baseline from Mankind", where "the baseline consist[s] in a condition that is (in some sense) normal for mankind" (Holtug, 2002, 369). Neither of them explicitly spells out the account, but I take it that the intended view is this:

MCA: An event *e* harms a subject *s* if, and only if, *e* makes *s* worse off (in terms of well-being) than the normal well-being level of mankind.

Again, I let HP + MCA denote the conjunction of HP and MCA. Holtug finds this position fatally flawed, because someone "may harm a person by removing some of his welfare, even if he is still better off than the norm" (2002, 369). HP + MCA counterintuitively implies that the state cannot justifiably intervene to prevent an event that radically lowers a subject's well-being level, if her level remains above the normal level of mankind.

Since HP is implausible combined with the mentioned comparative views, Holtug and Petersen suggest that the best candidate at least of the comparative kind is the Counterfactual Comparative Account (CCA) of harm.

CCA: An event *e* harms a subject *s* if, and only if, *s* would have been better off (in terms of welfare over her lifetime) in the absence of *e*.

On this account, harms are events that make a subject worse off than she otherwise would have been. It is not a surprise that the critics focus on this view, since it is the standard view of harm (among contemporary proponents are Boonin, 2008; Feit, 2019; Klocksien, 2012, 2019; Purshouse, 2016; and Purves, 2019).

In contrast to HP + TCA, HP + CCA does not rule out that the state can justifiably intervene against preventive harms, since they leave their subject worse off than they otherwise would have been. HP + CCA seems better positioned to handle cases where someone is born with a bad condition too. As long as a child's bad condition makes her worse off than she would otherwise have been, the child is harmed on CCA. Since norms of welfare levels are irrelevant on CCA, HP + CCA (in contrast to HP + MCA) does not rule out that the state can justifiably intervene to prevent an event that radically lowers a subject's well-being level, which is above the normal level of mankind.

<sup>3</sup>Jonas does not explicitly use the term "the Harm Principle", but she discusses a principle saying that harm prevention is necessary for state intervention.

However, the critics argue that HP + CCA too has fatal problems. According to Holtug and Petersen, HP + CCA implausibly fails to rule out that the state can justifiably intervene with failures to benefit. Examples of such failures to benefit are failing to invite someone else to a cruise (Holtug, 2002) or donating 5% of your salary instead of 6% (Petersen, 2014). If we assume that there is some event (e.g., a person deciding not to invite someone else on a cruise, or changing her mind about donating 6% of her salary and instead donating 5%) such that a subject is worse off than she would have been in the absence of the event, then CCA implies that the event harms the subject. According to Holtug and Petersen, HP *should* rule out that the state can justifiably intervene with such acts. But HP + CCA fails to do so. Holtug categorizes the above failures to benefit as one type of counterexample to HP, namely a type of counterexample that “consists in showing that the principle is compatible with [state intervention] when it should in fact rule it out” (Holtug, 2002, n. 16).

In response to problematic cases of failing to benefit, some suggest a moralized revision of the principle. (Some also think that a moralized principle is closer to Mill’s own view (Holtug, 2002, n. 37; Kleinig, 1978, 27), but I will not discuss this question of interpretation.) Petersen (2014) considers a revised principle saying that preventing *morally wrongful* harm to others is necessary for justified state intervention.<sup>4</sup> This is the Moralized Harm Principle (MHP):

MHP: If state intervention against an individual is justified, then this is at least partly because it prevents morally wrongful harm.

MHP + CCA is immune to the objection above, Petersen thinks, since it is not morally wrong to fail to invite someone on a cruise or to donate 5% of your salary instead of 6%. If so, MHP + CCA says that the state cannot justifiably intervene against these acts.

However, this response is seriously problematic nevertheless, Petersen thinks, because combining CCA with MHP makes MHP’s appeal to harm superfluous. To understand this argument, it is crucial to know that Petersen controversially claims that CCA implies that practically all acts and omissions are harmful. Focusing on state intervention in the form of criminalization, Petersen writes:

But if all acts and omissions are harmful, and if it is only morally wrongful harms that we have a reason to prohibit, then it is unclear what work theories of criminalization (at least when it comes to the part of the theory that entails harm as a moral factor for criminalization), which fit under the umbrella of a Harm View, actually do. (2014, 212)

According to Petersen, resorting to MHP + CCA makes it evident that the harm component in MHP does no work at all. This, Petersen thinks, gives us “reason to reject the harm principle and instead use a full-blown moral theory to justify which types of acts should be criminalized” (2014, 213). However, Petersen reasons, since the issue stems from CCA’s implication that practically all acts and omissions are harmful, an alternative is to reject CCA. According to Peterson it is not easy to come up with an alternative account to CCA which does not face the issues previously discussed in this section. Since both accepting and rejecting CCA thus leads to problems, Petersen concludes, proponents of a view like HP face a dilemma: “Either they can accept the counterfactual

<sup>4</sup>See also Holtug’s discussion of a moralized version of the principle (2002, 378).

baseline but then, they can do without the harm principle. Or they can reject the counterfactual baseline – but then they will have to formulate an alternative baseline which ... is no easy task” (Petersen, 2014, 199)

Jonas (2016) presents another argument against HP + CCA. By appealing to HP one can, for instance, argue that parents can rightfully lose custody of their child only if this prevents harm to the child. According to Jonas, however, HP + CCA is implausible in this context.

Sugar may look harmless compared to a diet high in heavy metals and saturated fat, or to near-starvation. Compared to a worthy regime of brown rice, fresh fruit and high quality proteins it may fare much the worse. The choice of counterfactual will determine whether a childhood experience is deemed harmful, but it is often unclear what should be chosen. (2016, 295)

According to Jonas, on CCA a given act or activity can be deemed harmful or harmless depending on what comparison we make. Putting a child on a diet high in sugar is thus harmful compared to a healthy diet, she reasons, but not compared to near-starvation. Since “counterfactual comparators lack a steady grounding”, Jonas thinks, we should reject HP + CCA.

“In conclusion,” Holtug says “none of the [accounts] considered gives rise to a plausible version of the Harm Principle” (Holtug, 2002, 373). If the critics are right, there is no plausible version of HP, no matter which of the accounts of harm is plugged into it.<sup>5</sup> According to the critics, HP + TCA and HP + MCA are clearly not extensionally adequate. The best candidate, at least out of the comparative accounts of harm, for being plugged into HP is allegedly CCA. However, the critics’ arguments conclude that also HP + CCA has fatal problems – for example, the problem with failures to benefit, which cannot be fixed by resorting to a moralized principle.

#### 4. In defense of HP

The previous section makes clear that the critics find HP implausible, no matter which of the discussed theories of harm is plugged into it. This type of criticism is potentially devastating for anyone embracing or relying on HP. But as of now, a systematic engagement with the arguments underlying the criticism is still lacking in the literature.

We earlier saw various counterexamples against HP in combination with (more and less) prominent accounts of harm. One may think that we therefore can conclude that HP is false. But that is too quick. Showing that HP is false takes more than presenting counterexamples against HP in combination with a set of accounts of harm. A counterexample against HP + a particular account of harm shows that either HP is false, or that account of harm is false, or both are false. But since the account of harm may be the only faulty conjunct, such a counterexample does not show that HP is false. A valid argument concluding that HP is false, which expresses the critics’ views, must therefore also include the premise that one of the accounts of harm is true:

<sup>5</sup>How to interpret Jonas’ (2016) conclusion is not clear. Although she argues against (for instance) the positions I call HP+CCA and HP+TCA, she does not conclude that HP necessarily is implausible. She thinks that different baselines (including the counterfactual baseline) should be used with HP in different contexts, depending on their particular normative import. I thank an anonymous referee for bringing this up.

- P1. Either TCA, MCA or CCA is true.<sup>6</sup>  
 P2. HP + TCA is false.  
 P3. HP + MCA is false.  
 P4. HP + CCA is false.  
 C1. HP is false.

On this argument, the conjunctions of HP and the alternative accounts of harm are false. And according to P1 one of these accounts is true, which means that HP must be false. I will mainly discuss the plausibility of P4 and P1. But let us first quickly consider P2 and P3. We saw that, according to the critics, preventive harms and cases where an individual is born with a bad condition constitute counterexamples to HP + TCA. If they are right, then P2 is true. Moreover, the critics find the following a counterexample to HP + MCA: an event that does not make an individual worse off than the average of mankind, but radically decreases the individual's wellbeing. If this is right, then P3 is true too. I do not oppose the critics on these matters; I grant that P2 and P3 are true.

Let us move on to P4. We might expect that the critics have convincingly supported P4 by giving clear-cut counterexamples to HP + CCA, simply because they mainly focus on HP + CCA. I think that there are such counterexamples, but surprisingly they are not acknowledged by the critics. The critics argue that HP + CCA is implausible because it fails to rule out that the state justifiably can intervene against individuals who simply fail to benefit others. However, the relevant cases of failing to benefit are not counterexamples to HP + CCA. This is because, on HP, harm prevention is as mentioned not a sufficient condition for justified state intervention, and harm prevention may not be the entire reason why the state can justifiably intervene (when it can do so). This means that HP + CCA proponents can point to other – non-harm related – factors to explain why the state cannot justifiably intervene with someone who, for instance, fails to invite someone else to a cruise. They might claim, say, that such intervention would be a highly inefficient use of state resources. So HP + CCA proponents can coherently say both that the relevant failures to benefit are harmful and that the state cannot justifiably intervene against them. Holtug, seemingly aware of this, insists: “my guess is that typical proponents of [HP] will want this principle to rule out coercion here; after all, by coercing these innocent people, the state [does not] prevent harm” (2002, n. 16). Even if this correctly represents the intention of typical HP proponents, it does not mean that the failures to benefit are counterexamples to HP + CCA. Thus, the relevant cases of failure to benefit do not support P4.

Because we have yet to see counterexamples to HP + CCA, one might think that P4 is false (which means that HP + CCA is true). If that were right, my main objective of defending HP would be achieved. But I do not think that HP + CCA is defensible. This is partly because HP + CCA implausibly excludes justified state intervention in some preemption cases.<sup>7</sup>

Consider a case where someone robs a homeless person of her few belongings, causing her much distress and hunger. The robber has an accomplice who would have done

<sup>6</sup>Although some other accounts of harm are discussed by the critics, I have omitted these accounts here since the critics seem to think that they are implausible.

<sup>7</sup>Petersen (2014) discusses a preemption case formulated by Norcross (2005). However, it is not under the description of “preemption”, and his idea of what problem the case poses for CCA differs from mine. The preemption problem, as well as the similar overdetermination problem, for CCA is well discussed in the debate about the nature of harm. See for example Bradley (2012), Hanna (2016), Johansson & Risberg (2019), Kloksiem (2012), Northcott (2015), and Rabenberg (2015).

the same thing if the robber had not. This is called a “preemption” case, since the accomplice’s action is preempted. The cause of the homeless person’s distress and hunger is the actual robber’s actions. HP + CCA counterintuitively excludes that the state can justifiably stop the robber’s actions. This is because the homeless person would not have been better off had the robber acted differently. The same effect would have been brought about by the accomplice.

A natural response might be that the state can justifiably stop *both the actual robber and the accomplice* from robbing the homeless person. Stopping both of them, HP + CCA defenders can argue, does prevent a harmful event. The harmful event could be that the homeless person loses her belongings. As she is better off in the absence of losing her belongings, we assume, CCA implies that she is harmed by that event.

This response, however, is not convincing. First, I doubt that this response works for HP + CCA generally. This picture fits poorly with the origin-centered version of HP, presented in section 2, and seems restricted to proponents of origin-neutral versions. For even if the state can prevent an event that CCA counts as harmful (the homeless person losing her belongings), it is not clear that either the robber or the accomplice is the origin of that harm. As we saw, the robber’s act is harmless on CCA. It is hard to see how the harm could originate from the robber despite the fact that the robber’s act is harmless. It is doubtful that a harmful event can originate, in any plausible sense of the term, from an agent whose actions are harmless.

Second, this response is unsatisfactory regardless of which version of HP one assumes. Let us say that HP + CCA does not rule out that the state can justifiably stop both the robber and the accomplice, either because the above issue for origin-centered versions of HP can be solved or because an origin-neutral version is true. Even so, the problem for HP + CCA remains. Imagine that the state *cannot*, or *will not*, intervene against both the robber and the accomplice. If so, then HP + CCA counterintuitively rules out that the state can justifiably stop the robber only. Stopping only the robber means that the accomplice steps in and steals the homeless person’s belongings. Since the homeless person is not better off in that scenario, the state has not prevented a harmful event (on CCA). Since HP + CCA counterintuitively rules out that the state justifiably can stop the robber in a case where the state cannot, or will not, also stop the accomplice, HP + CCA remains extensionally inadequate.<sup>8</sup>

So far I have not challenged the truth of any premise in the argument. I granted that we should accept P2 and P3 and argued that there are good reasons to accept P4, despite the critics’ failure to acknowledge them. Let us now discuss P1. This premise says that one of the accounts of harm (TCA, MCA or CCA) is true. I think that this premise is key and that we should reject it.

Perhaps the most obvious way to question P1 is to appeal to a certain account of harm *other than* TCA, MCA or CCA. Although CCA is the most commonly accepted account in the debate, there are contenders beyond TCA and MCA. Since not all accounts have been dismissed as viable candidates for being plugged into HP, one might think, there can be an account of harm such that HP, when that account is plugged into it, does not face counterexamples.

One might, for instance, advocate a non-comparative account of harm (see for example Harman, 2009; Shiffrin, 1999). Here is a common version of the non-comparative account (NC): *an event e harms a subject s iff e causes s to be in an*

<sup>8</sup>One can easily construe *overdetermination* cases, which HP+CCA will similarly struggle with. For a discussion of the overdetermination problem for CCA, see for instance Shiffrin (2012) and Tadros (2014).



*intrinsically bad state*. HP + NC is better suited to handle the preemption case discussed above, since the robbery plausibly causes some intrinsically bad state. After all the victim suffers from distress and hunger. Thus, in contrast to HP + CCA, HP + NC does not rule out that the state can justifiably intervene against such acts.

However, HP + NC faces other issues. HP + NC implausibly excludes justified state intervention against acts that cause someone to be worse off, but not so badly off that they are in a bad state.<sup>9</sup> Imagine that an organization attacks a city's water supply system, which results in slightly poorer health for part of the city. Everyone in this part of the city is quite well off and the contaminated water does not cause anyone to be in an intrinsically bad state. But their wellbeing levels decrease. Since no one is caused to be in an intrinsically bad state, HP + NC implausibly excludes that the state can justifiably intervene with the organization attacking the water supply system.

In general, HP seems problematic in combination with any *undergenerating* account of harm, that is, an account that fails to identify the harm in some harmful cases. For instance, NC undergenerates harm in cases like the attack on the water supply; and CCA undergenerates harm in the preemption case. If we combine an undergenerating account of harm with HP, the principle unavoidably excludes the possibility that the state can justifiably intervene in some harmful cases. Crucially, there is some reason to suspect that all (or most) suggested accounts of harm undergenerate harm of some kind. Bradley (2012) discusses various accounts and concludes that all accounts have such serious issues that the harm concept is unfit for moral theorizing altogether. In an examination of various causal accounts, Carlson et al. (2021a) find all of them wanting. If these skeptical conclusions are right, then HP is likely to be problematic in combination with all (or most) suggested theories of harm. Therefore, it is difficult to refute this type of criticism by pointing to another account of harm, which in conjunction with HP faces no counterexample. I know of no such account.

Fortunately for friends of HP, there is another way to question P1. That is to show that the problems for the conjunctions of HP and the accounts of harm are inherited from the respective accounts of harm. The problematic conjunct is thus the account of harm and not HP. On a closer look at the critics' arguments, they reasonably do not shed doubt on HP. The main target of the critics' arguments is (a version of) HP + CCA, and they should, if anything, make us doubt CCA.

As we saw, Holtug is sympathetic to the idea that HP should be combined with CCA, but thinks that this position cannot offer sufficient protection of liberty, as it fails to rule out justified intervention in the previously discussed cases of failing to benefit. This alleged problem for HP + CCA is clearly inherited from CCA's failure to accommodate the relevant cases of failures to benefit. In fact, Holtug's own words indicate this. Regarding state intervention against the agents who fail to benefit, he writes: "by coercing these innocent people, the state [does not] prevent harm" (2002, n. 16). In other words, these failures to benefit are, contrary to CCA's implication, not harms. CCA's counterintuitive implication that they are harmful is a recognized problem for CCA; and a satisfying solution seems to be lacking.<sup>10</sup> What is important in this context is

<sup>9</sup>This problem for non-comparative accounts is also called the "worse-off-but-not-badly-off" problem (Carlson et al., 2021a; Hanser, 2008; Rabenberg, 2015). Holtug (2002) mentions a similar problem. But he frames that as a problem for "quantitative views", rather than a problem for non-comparative views.

<sup>10</sup>For discussion of the failure to benefit problem, see Bradley (2012), Feit (2019), Folland (Manuscript), Johansson & Risberg (2020), and Purves (2019).

that these failures to benefit are, if anything, problematic for CCA. The problem is plausibly *not* that HP prescribes harm a normative role that harm is unable to play.

Second, despite CCA's popularity as an account about the nature of harm, it is a rather curious idea that HP is plausibly combined with CCA. According to a prominent objection, CCA is incompatible with commonly accepted, and plausible, views about the normative significance of harm (Carlson, 2019; Carlson et al., 2021b). Several of CCA's proponents respond by claiming that philosophers tend to overstate harm's normative significance. For instance, in light of the failures to benefit problem for CCA, Neil Feit (2019) argues that some harms lack moral significance altogether. The idea that harm can lack normative significance does not go well with HP. If CCA is true, and if harm can fail to be normatively significant, it is implausible that harm prevention always is part of the justification when the state justifiably intervenes against an individual. Differently put, if harm can be morally insignificant it plausibly cannot play the justificatory role that HP assigns it. The idea that CCA might successfully be combined with HP is therefore peculiar.

Moving on, as we saw, Petersen says that although MHP (the Moralized Harm Principle)+CCA seems plausible, the principle's appeal to harm is actually superfluous. He thinks that if only *morally wrongful* harms are candidates for justifying intervention, then the justificatory work "– at least as far as the element of harm is concerned – could be done entirely with reference to an independent theory of moral wrongfulness" (2014, 202). The reason why the harm component in MHP + CCA does no work at all, Petersen thinks, is that practically all acts and omissions are harmful on CCA.

However, and contrary to Petersen's assertion, CCA does not imply that all acts and omissions are harmful. For one thing, all *acts* do not make someone worse off than she otherwise would have been. That I lift my finger from the keyboard is, in most situations, not harmful on CCA. It is harmful on CCA only in rare cases where had I not lifted my finger from the keyboard I would have, say, donated money or paid someone a much-needed compliment. In most situations, an act like lifting my finger does not satisfy CCA's condition, since no subject would have been better off had I not lifted my finger. Furthermore, not all *omissions* are harmful on CCA. For example, we can correctly describe the mere absence of my drinking a glass of water as an omission: I omit to drink a glass of water. It is not the case that mere absences of events are harmful on CCA. For the theory is about the harmfulness of *events*; and it therefore has implications only about events. CCA has no implications about my omitting to drink a glass of water or similar cases, since they involve no relevant event. As CCA does not imply that all omissions and acts are harmful, I do not think that Petersen's argument threatens either proponents of any version of HP or, for that matter, proponents of CCA.<sup>11</sup>

As we saw, Jonas thinks that HP + CCA is problematic because "the choice of counterfactual will determine" whether some event is harmful; "but it is often unclear what should be chosen" (Jonas, 2016, 295). Of course, it is not *simply for us to decide* what the relevant counterfactual comparison is on CCA. CCA specifies that the counterfactual comparison is the scenario that would have happened in the absence of the relevant event. However, it can be hard to know what would have happened in absence of an

<sup>11</sup>There are further issues with Petersen's reasoning. His argument is unsound *even if* CCA had this implication. The claim that all acts and omissions are harmful is counterintuitive, I think, to the extent that a theory cannot be true if it implies that. Thus, if CCA implies that – and even my lifting a finger under normal circumstances is harmful – then the idea that a principle like HP should be combined with CCA is futile.

event. This suggests that there might be practical issues with using HP + CCA. One might think that it is too hard for decision-makers, lawyers, etc. to determine in practice which cases are harmful and thereby relevant for HP + CCA. But this objection against HP + CCA is questionable, mainly because it sounds reasonable that determining harm is not a straightforward business. We should expect, I think, that we do not always know if a particular event is harmful or not. Moreover, although our epistemic limitations are regrettable, it is a practical issue how decision-makers and lawyers make good decisions in practice. This practical issue should be distinguished from (although it perhaps relates to) both the issue of what harm is and what the relationship is between harm prevention and justified state intervention. Importantly, however, if this is a problem for HP + CCA, it undeniably stems from CCA. If the uncertainty regards whether a subject would have been better off in the absence of an event, then the uncertainty stems from CCA and not from HP. In any case, the main issue for HP + CCA is not, I think, unstable counterfactual comparison states. Rather, it is the counterintuitive implications previously spelled out in this section.

Finally, a possible reaction is that I dismiss CCA, or other accounts of harm, too easily. Perhaps one hesitates to assign so much weight to judgments about particular cases as to conclude that we should dismiss the discussed accounts of harm. This would make a difference for the overall argumentation, because if we should not take these judgments seriously (or as seriously as to dismiss the accounts), then P1 might be true. Even if there *appear* to be counterexamples to all the accounts, one can argue, one account is true. I take no stand on how much weight intuitive judgments ought to be assigned. Regardless, the criticism against HP fails even if the judgments carry less or even minimal weight. For if the judgments should not be taken seriously when analyzing the plausibility of the accounts of harm, then they should reasonably not be taken seriously when analyzing the plausibility of HP in conjunction with some account of harm. And if the judgments are insufficiently weighty for disproving HP in conjunction with the accounts of harm, then there is no support for P2, P3 and P4. Thus, supporting P1 by arguing that the judgments about the particular cases should not carry much weight undermines the support for the other premises.

If I am right so far, this type of criticism fails to show that HP is false. There might be many reasons to doubt HP in conjunction with various accounts of harm, some of which are discussed here. But we have yet to see a good reason to think that HP is the problematic conjunct. This is naturally good news for friends of HP. However, some worries remain. One might wonder: given that there seems to be no problem-free conjunction of HP + a theory of harm, how could HP be plausible?

## 5. Must proponents of HP refer to a theory of harm?

HP is only one of many normative principles that appeal to the concept of harm. For instance, many formulations of the doctrines of double effect (Foot, 1967; Quinn, 1989) and doing and allowing (Woollard, 2015) appeal to the concept of harm. There are various principles about harm's moral relevance. A common principle among them says that *it is pro tanto wrong to harm*. According to Russ Shafer-Landau, "[t]he default position in ethics is this: do no harm. It is sometimes morally acceptable to harm others, but there must be an excellent reason for doing so or else the harmful behavior is unjustified" (2021, 6). The debate about the so-called non-identity problem hosts the *no harm, no foul* principle (Boonin, 2014, 5). A simple formulation of this principle is that conduct is morally wrong only if it is harmful. An even stronger principle,

combining the previous principles, says that “[c]onduct is [pro tanto] morally wrong if and only if it is harmful to others” (Brown, 2010, 126).

Most commonly, philosophers do not explicitly discuss whether the nature of *harm* is relevant for debates about these principles. However, a couple of philosophers indicate their stance on this matter. One example is Quinn’s (1989) discussion (and defense) of the Doctrine of Double Effect (DDE). This principle roughly says that in pursuing the same good outcome, an intended harmful effect is harder to morally justify than a merely foreseen harmful effect. He clarifies that: “[h]arm is meant in a very broad sense that includes the loss of life, rightful property, privacy, and so on. In my examples, the relevant harm will usually be the loss of life” (1989, n. 1). Quinn goes on to discuss DDE without referring to, or commenting on, any account of harm. Quinn apparently finds giving examples of some core harms sufficient. Another example is Woollard’s (2015) more explicit approach, which appears in her defense of the Doctrine of Doing and Allowing (DDA). On this doctrine, doing harm is harder to morally justify than merely allowing harm. After noting that the notion of harm is controversial, Woollard remains agnostic about what the correct account of harm is. She thinks that for the purposes of discussing the plausibility of DDA, one need not go into the question of what harm is. Rather, she claims that “it is enough that we can recognize paradigm cases of harm” (2015, 18).

In contrast to these approaches, the critics of HP demand much more of HP proponents. Clearly, the critics rest their arguments on the assumption that HP is plausible only if there is a full-blown, problem-free (at least to a high degree) theory of harm that we can refer to. In light of alternative approaches, it is far from obvious that we should assume that proponents of HP need a full-blown, problem-free theory of harm to refer to. One might worry, though, that without a single, full-blown, generally accepted, account of harm we cannot use the notion of harm in HP or other similar principles. However, I suggest that we can.

There clearly is a set of judgments, or intuitions, about harms and harmful cases, which philosophers widely agree about. Among them are the judgments that underlie arguments for and against accounts of harm earlier discussed: that the act of robbery is harmful although its negative effects would have occurred even if the robber had not performed the act, that simply failing to invite someone to a cruise is not harmful, etc. Furthermore, judgements of this kind have in several instances led proponents of accounts of harm to revise their account. For instance, proponents of CCA have substantially modified CCA to accommodate judgements about preemption and overdetermination cases (Feit, 2015; Norcross, 2005), failing to benefit cases (Purves, 2019) and group harm (Norcross, 2005). These observations show that there is a core of the harm concept, or a core set of judgments about harm, about which philosophers widely agree. The upshot is that we do seem to know enough about harm to use the notion of harm. Reasonably, we can maintain that our well-considered judgments about harm are plausible, while the debate about the nature of harm continues.

Naturally, there is also *disagreement* about whether some classes of cases are harmful or not. One such class is that of *non-identity cases* (Boonin, 2014; Parfit, 1984). A non-identity case typically involves an act that causes a subject to exist with a low, but positive, wellbeing level. Although the subject’s existence is conditioned on that act, and the subject’s life is worth living, it seems that such acts are sometimes morally wrong. Importantly, whereas many agree that agents act morally wrongly in some non-identity cases, it is controversial whether the subjects are *harmed* (Boonin, 2008; Gardner, 2015; Harman, 2004; Parfit, 1984).

Although the potential relevance of disagreement cannot be exhaustively discussed here, it is worth noting that disagreement of this kind is not necessarily problematic for HP proponents. Whether this sort of disagreement is problematic depends on whether state intervention in these cases is plausibly justified. In many of the cases where there is disagreement about harmfulness, non-identity cases included, it is likely to be controversial whether state intervention is justified. If so, a plausible diagnosis is that it is controversial whether the state can justifiably intervene, *because* it is unclear whether such cases involve harm. This would speak in favor of the idea that harm is normatively important in the way that HP states. Hence, the fact that there is disagreement about some classes of cases, with regard to their harmfulness, could support HP, rather than oppose it.

Not only is the demand that HP needs to refer to a full-blown theory of harm, in my view, uncalled for. What is more, accepting a strong, extensive, demand in this spirit also risks leading to a more widespread skepticism. Using the kind of reasoning that the critics do, it is likely that focusing on what, for instance, *prevention* is also leads to skepticism. Since HP invokes the concept of prevention, critics might demand that proponents of this principle must refer to a theory about what prevention is. Critics can argue that, as it turns out, there is no problem-free theory about the nature of prevention that can be plugged into this principle, and the principle should therefore be rejected. This indicates that it is unreasonably demanding to require that there is a full-blown, and problem-free, theory of harm that HP proponents can refer to. Moreover, the skepticism plausibly generalizes to all principles that appeal to concepts that are subject to competing understandings. Some examples are principles appealing to knowledge, time, art, personal identity, or moral rightness. It is hard to see how widespread skepticism can be avoided, if we demand that principles like HP need to refer to full-blown, problem-free theories of concepts that they appeal to.

One might of course think that, in contrast to other principles that appeal to various concepts, the concept of harm is *central* to HP. The demand on HP proponents to say what harm is, one might think, is not analogous to a demand on proponents of any principle that appeals to a controversial concept. But rather the demand on HP proponents might be analogous to that on, for instance, hedonistic utilitarians to say what pleasure and pain is.<sup>12</sup> And as the latter plausibly are required to refer to a full-blown, plausible account of pleasure and pain, one might argue, the former are required to refer to a full-blown, plausible account of harm.

I lack space here to fully examine this potential analogy, but there are reasons to think that my sketched view is plausible despite these considerations. First, if the analogy holds, my view compels me to say that the hedonistic utilitarian need not refer to a full-blown, problem-free account of pleasure and pain. But this is not necessarily troublesome if there is a set of core judgments about pleasure and pain (as, I argued, there is a set of core judgments about harm), which the hedonistic utilitarian can utilize. So the corresponding view about the demand on hedonistic utilitarianism is not necessarily problematic. Second, if this is a genuine issue for hedonistic utilitarians, then it plausibly targets its competitors too. Abandoning hedonistic utilitarianism for, say, some version of rights theory makes you obligated to spell out a full-blown and problem-free account of rights. If that is neither harder nor easier than spelling out

<sup>12</sup>Anscombe and (building on her work) Rawls judged “the concept [pleasure] so obscure and problematic that theories placing weight upon it, such as hedonistic utilitarianism, should be rejected out of hand” (Katz, 2016, sec. 1.2).

what pleasure and pain is, it cannot be a blow either to hedonistic utilitarianism or to the rights theory, in a comparison of their plausibility. Thus, I think that these considerations do not render my sketched view implausible.

## 6. Concluding remarks

Critics argue that HP is implausible, due to difficulties with finding an account of harm that can be plugged into it. No account of harm can be plugged into the principle, they argue, such that it amounts to a plausible version of the principle. After examining the criticism's strongest form, I find that it shows neither that HP is false nor that it is useless. Moreover, the criticism relies on the assumption that HP proponents must be able to refer to a full-blown, problem-free, theory of harm that can be plugged into HP. We should reject that idea, I argue, partly because if we accept it then widespread skepticism threatens, not only about various principles that appeal to harm but about many kinds of principles. It is fair to conclude, I think, that the demand on HP proponents to explicate what harm is, plausibly is not strong, or extensive, enough to support the criticism against HP.<sup>13</sup>

**Competing interests.** The author declares none.

## References

- Bell, Melinda Constantine.** 2020. John Stuart Mill's Harm Principle and Free Speech: Expanding the Notion of Harm, *Utilitas*, **33**: 1–18.
- Boonin, David.** 2008. How to Solve the Non-Identity Problem, *Public Affairs Quarterly*, **22**: 129–59.
- Boonin, David.** 2014. *The Non-Identity Problem and the Ethics of Future People* (New York: Oxford University Press).
- Bradley, Ben.** 2012. Doing Away with Harm, *Philosophy and Phenomenological Research*, **85**: 390–412.
- Brink, David O.** 2013. *Mill's Progressive Principles* (Oxford: Oxford University Press).
- Brown, Donald G.** 2010. Mill on the Harm in Not Voting, *Utilitas*, **22**: 126–33.
- Browne, Alister.** 2016. Mill on Mental Health Acts, *Utilitas*, **28**: 1–18.
- Carlson, Erik.** 2019. More Problems for the Counterfactual Comparative Account of Harm and Benefit, *Ethical Theory and Moral Practice*, **22**: 795–807.
- Carlson, E., J. Johansson, and O. Risberg.** 2021a. Causal Accounts of Harming, *Pacific Philosophical Quarterly*. Online first. <<https://doi.org/10.1111/papq.12390>>
- Carlson, E., J. Johansson, and O. Risberg.** 2021b. Well-Being Counterfactualist Accounts of Harm and Benefit, *Australasian Journal of Philosophy*, **99**: 164–74.
- Dripps, Donald A.** 1998. The Liberal Critique of the Harm Principle, *Criminal Justice Ethics*, **17**: 3–18.
- Dworkin, Gerald.** 1972. Paternalism, *The Monist*, **56**: 64–84.
- Edwards, James.** 2014. Harm Principles, *Legal Theory*, **20**: 253–85.
- Feit, Neil.** 2015. Plural Harm, *Philosophy and Phenomenological Research*, **90**: 361–88.
- Feit, Neil.** 2019. Harming by Failing to Benefit, *Ethical Theory and Moral Practice*, **22**: 809–23.
- Folland, Anna.** Manuscript. Feit on the Normative Importance of Harm.
- Foot, Philippa.** 1967. The Problem of Abortion and the Doctrine of Double Effect, *Oxford Review*, **5**: 5–15.
- Gardner, Molly.** 2015. A Harm-Based Solution to the Non-Identity Problem, *Ergo*, **2**: 427–44.
- Hanna, Nathan.** 2016. Harm: Omission, Preemption, Freedom, *Philosophy and Phenomenological Research*, **93**: 251–73.

<sup>13</sup>I would like to thank Jens Johansson, Erik Carlson, Olle Risberg, and participants of the Philosophy, Politics and Economics Seminar at the Institute for Future Studies and the Higher Seminar in Practical Philosophy at Uppsala University for detailed and helpful feedback on earlier versions of this article. For valuable comments I am also grateful to Dale E. Miller and two anonymous referees for *Utilitas*.

- Hanser, Matthew.** 2008. The Metaphysics of Harm, *Philosophy and Phenomenological Research*, 77: 421–50.
- Harman, Elizabeth.** 2004. Can We Harm and Benefit in Creating? *Philosophical Perspectives*, 18: 89–113.
- Harman, Elizabeth.** 2009. Harming as Causing Harm, in *Harming Future Persons*, ed. by Melinda A. Roberts and David T. Wasserman (Dordrecht: Springer), pp. 137–54.
- Holtug, Nils.** 2002. The Harm Principle, *Ethical Theory and Moral Practice*, 5: 357–89.
- Johansson, J., and O. Risberg.** 2019. The Preemption Problem, *Philosophical Studies*, 176: 351–65.
- Johansson, J., and O. Risberg.** 2020. Harming and Failing to Benefit: A Reply to Purves, *Philosophical Studies*, 177: 1539–48.
- Jonas, Monique.** 2016. Assessing Baselines for Identifying Harm: Tricky Cases and Childhood, *Res Publica*, 22: 387–404.
- Katz, Leonard D.** 2016. Pleasure, in *The Stanford Encyclopedia of Philosophy* <<https://plato.stanford.edu/>> [accessed 11 June 2021].
- Kleinig, John.** 1978. Crime and the Concept of Harm, *American Philosophical Quarterly*, 15: 27–36.
- Klocksiehm, Justin.** 2012. A Defense of the Counterfactual Comparative Account of Harm, *American Philosophical Quarterly*, 49: 285–300.
- Klocksiehm, Justin.** 2019. The Counterfactual Comparative Account of Harm and Reasons for Action and Preference: Reply to Carlson, *Ethical Theory and Moral Practice*, 22: 673–77.
- Mill, John Stuart.** [1859] 1977. *On Liberty*, in *Collected Works of John Stuart Mill*, vol. XVIII, ed. by John M. Robson (Toronto: University of Toronto Press), pp. 213–310.
- Norcross, Alastair.** 2005. Harming in Context, *Philosophical Studies*, 123: 149–73.
- Northcott, Robert.** 2015. Harm and Causation, *Utilitas*, 27: 147–64.
- Parfit, Derek.** 1984. *Reasons and Persons* (Oxford: Clarendon Press).
- Persson, Ingmar.** 1998. Mills distinktion mellan frihet i första person och andra, *Tidskrift för politisk filosofi*, 3: 5–18.
- Petersen, Thomas Søbirk.** 2014. Being Worse Off: But in Comparison with What? On the Baseline Problem of Harm and the Harm Principle, *Res Publica*, 20: 199–214.
- Purshouse, Craig.** 2016. A Defence of the Counterfactual Account of Harm, *Bioethics*, 30: 251–59.
- Purves, Duncan.** 2019. Harming as Making Worse Off, *Philosophical Studies*, 176: 2629–56.
- Quinn, Warren S.** 1989. Actions, Intentions, and Consequences: The Doctrine of Double Effect, *Philosophy & Public Affairs*, 18: 334–51.
- Rabenberg, Michael.** 2015. Harm, *Journal of Ethics and Social Philosophy*, 8: 1–32.
- Saunders, Ben.** 2016. Reformulating Mill's Harm Principle, *Mind*, 125: 1005–32.
- Shafer-Landau, Russ.** 2021. *The Fundamentals of Ethics*, 5th ed. (New York: Oxford University Press).
- Shiffrin, Seana Valentine.** 1999. Wrongful Life, Procreative Responsibility, and the Significance of Harm, *Legal Theory*, 5: 117–48.
- Shiffrin, Seana Valentine.** 2012. Harm and its Moral Significance, *Legal Theory*, 18: 357–98.
- Tadros, Victor.** 2014. What Might Have Been, in *Philosophical Foundations of the Law of Torts*, ed. by John Oberdiek, (Oxford: Oxford University Press), pp. 171–92.
- Turner, Piers Norris.** 2014. “Harm” and Mill's Harm Principle, *Ethics*, 124: 299–326.
- Westmoreland, Robert.** 2020. The Liberty of the Liberty Principle, *Res Publica*, 26: 337–55.
- Woollard, Fiona.** 2015. *Doing and Allowing Harm* (New York: Oxford University Press).