CAMBRIDGE
UNIVERSITY PRESS

## ARTICLES

# Some Acts Really Harm: A Defense of the Standard Account versus Norcross's Contextualism

Ben Eggleston ⓘD

Department of Philosophy, University of Kansas, Lawrence, KS, USA
Email: eggleston@ku.edu

**Abstract**

An important strand of argument in Alastair Norcross's *Morality by Degrees: Reasons without Demands* is the rejection of the standard account of harm, which underwrites non-comparative statements of the form "act A harms person X." According to Norcross, the correct account of harm is a contextualist one that only underwrites comparative statements of the form "act A results in a worse world for X than alternative act B, and a better world than alternative act C." This article criticizes Norcross's contextualist account and his rejection of the standard account. It follows that moral theorists of all kinds should not be deterred by Norcross's arguments from continuing to rely on the standard account and using it to non-comparatively categorize some acts as harmings.

**Keywords:** scalar consequentialism; contextualism; harm

## 1. Introduction

"First, do no harm," says the Hippocratic Oath. "No harm, no foul," has been generalized far beyond the basketball court. And one of the most famous principles in political philosophy is 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" (Mill 1859, p. 223).

These examples, partially echoing and building on a list given by Ben Bradley (Bradley 2012, p. 390), illustrate how the category of harm is one of the key regulative concepts of our interactions with other people. Most of us pay close attention to whether and when others harm us and our loved ones, and most of us also try, albeit with varying levels of earnestness, to avoid harming others. For both of these purposes, and others (such as tort law), it is critical that the question of whether a particular act harmed a particular person have a definite answer.

The denial of this possibility is one of the many striking theses that Alastair Norcross argues for in *Morality by Degrees: Reasons without Demands* (Norcross 2020). Specifically, Norcross claims that "There is no fundamental non-comparative moral fact of the form 'act A harms person X'" (p. 101). Instead, "The fundamental moral facts, as regards harm, are of the form 'act A results in a worse world for X than

alternative act B, and a better world than alternative act C'" (p. 101). The same goes for facts regarding benefit (pp. 107 and 119). In essence, no acts *really* harm (though, as we will see, Norcross does not discourage the use of the word 'harm' as long as it is interpreted in a certain way).

Norcross arrives at these positions after critiquing the standard account of what makes an act a harming, or a benefiting. In place of that account, Norcross proposes a contextualist account. In this article, I will raise some concerns about Norcross's contextualist account and defend the standard account against Norcross's critique. It will follow that fundamental facts of the form "act A harms person X" are, in many cases, perfectly secure after all – or, at least, not made less secure by Norcross's critique. There is, of course, considerable debate over rival accounts of harm, with defenders of the standard account facing objections on several fronts. Useful entry points into the current debate are provided by Ben Bradley (2012), Molly Gardner (2021), and Erik Carlson *et al.* (2022), though here I will limit my focus to Norcross's arguments and accept his use of the label 'standard' for the account in question. I will also follow Norcross in limiting my focus to "the notion of harm *all things considered*, as opposed to harm in some respect or other" (p. 84). If a child physically resists receiving a medically advisable vaccination shot but their parent uses moderate force to make them comply, the parent harms the child in some respect, but the action is (probably) not an instance of harm all things considered.

It should be noted that although the issue of whether some acts really harm arises within Norcross's arguments for a particular form of consequentialism (which I will address briefly below), this issue is of far broader interest. As Norcross observes, prohibitions on harming are standard features of deontological moral views (p. 15). In fact, an adherent of any ethical outlook – whether consequentialism or common-sense morality, Kantian deontology or Rossian pluralism, contractarianism or egoism – will want to be able to make judgments of the form "act A harms person X." Thus, a defense of the standard account should be of interest to moral theorists of all stripes, not just consequentialists.

## 2. Norcross's arguments

### 2.1. Norcross's negative agenda

Norcross's rejection of fundamental moral facts of the form "act A harms person X" can be understood as the third phase of a campaign to show that several kinds of claims that are normally taken to express fundamental moral facts do not actually do so. The first phase is to show that instead of categorizing acts as right or wrong, we should simply regard certain acts as better or worse than others, based on their consequences. This is the signature thesis of Norcross's scalar consequentialism, and he says it is the "focus" of the book (p. 11). In brief, his argument for this thesis is that it does not make sense for consequentialists to draw "all-or-nothing" (p. 22) distinctions between some acts and others, such as the "right vs. wrong" distinction. This is because, quite often, a person's possible acts will lead to states of affairs that do not differ from one another in an all-or-nothing way; rather, the possible outcomes differ gradually, as in a case where a person can give various possible amounts of money to charity (pp. 22–24). Since consequentialists hold that the moral quality of acts is determined entirely by the moral quality of their outcomes, their evaluations of acts should also be gradual, rather than all-or-nothing. On this basis, Norcross writes the following. (Although he puts this claim in terms of utilitarianism, he intends for it to apply to all consequentialist theories.)

My suggestion is that utilitarianism should be treated simply as a theory of the goodness of states of affairs and of the comparative value of actions (and, indeed, of anything appropriately related to states of affairs, such as character traits, political systems, pension schemes, etc.), which rates possible alternatives in comparison with each other. This system of evaluation yields information about which alternatives are better than which and by how much. (p. 27)

This statement of Norcross's view explains the title of the book: morality is a matter of degree (rather than being all-or-nothing). As for the subtitle, although the theory presented yields reasons for choosing some acts rather than others, it does not issue demands (p. 21), because it does not demand that agents perform any acts, because it does not declare any acts to be wrong.

The second phase of Norcross's campaign against several kinds of putative fundamental moral facts regards the distinction between good and bad actions. It might appear that a consequentialist should be able to say what makes actions qualify as good, probably by saying that the goodness of an action is a function of the goodness of its consequences (p. 49). In particular, "It would appear that a consequentialist can classify as good any action that will *on balance, bring about more good than harm*, or whose *consequences are, on balance, good*, or whose *tendency to augment the good of the community is greater than any it has to diminish it*" (pp. 54–55). Norcross argues, however, that despite appearances, this proposal cannot be cashed out in any satisfactory way. It would be beyond the scope of my purpose in this article to discuss all of Norcross's remarks on this topic, but a simple example will convey the flavor of his approach. Consider the proposal that an action is good if the overall level of welfare (of all those affected by the action) is higher after it is performed than it was before, and bad if the overall level of welfare is lower (p. 55). This might seem plausible enough. But consider Norcross's case of *Doctor* (pp. 55–56), in which the following is true: a patient is terminally ill and their illness is causing them more and more suffering; the best treatment also causes suffering, but less suffering than the patient would otherwise experience; a doctor administers this treatment. The proposal under consideration would say the doctor's action was bad, but this is clearly incorrect (p. 56). The obvious next move is to say that the action was good because it made the patient better off than they would have been otherwise, counterfactually, even if not better off than before, temporally (p. 56), and I will comment on this idea further in relation to harm and benefit, in the next section. Norcross explores and rejects many other possible criteria for good and bad actions, concluding that "just as consequentialists should abandon (at least at the fundamental level) the notions of right and wrong actions … consequentialists should also abandon the notions of good and bad actions" (p. 74).

As I mentioned above, Norcross's discussion of harm is the third phase of his campaign against several kinds of putative fundamental moral facts. I turn to this discussion now.

## 2.2. Norcross's critique of the standard account of harm

Norcross's argument proceeds in stages, working from a (knowingly) naive starting point to a more sophisticated view. The naive starting point is the thought that an act harms a person if they are worse off after the act than they were before (p. 82). This view intentionally echoes the simple-minded proposal about good and bad actions discussed above, which was easily defeated by the case of *Doctor*. Norcross points out that *Doctor* defeats this view about harm in the same way. Specifically, this view implies

that the doctor has harmed the patient, since the patient is worse off than before. But clearly the doctor benefitted the patient; therefore, this view is wrong (p. 83).

To improve upon this naive view, Norcross presents the following principle, stating the standard account of harm:

> HARM: An act A harms a person P iff P is worse off, as a consequence of A, than she would have been if A hadn't been performed. An act A benefits a person P iff P is better off, as a consequence of A, than she would have been if A hadn't been performed. (p. 83)

Norcross notes that "At first glance, this seems pretty straightforward" (p. 83). If tenable, this principle would provide a criterion for fundamental facts of the form "act A harms person X." However, Norcross argues this principle is not tenable. His strategy is to point out that the principle can be read in multiple ways, corresponding to multiple possible interpretations of the clauses referring to the non-performance of the act, and to argue that each resulting version of the principle is unsatisfactory.

The first interpretation of the non-performance clauses that Norcross considers is that the agent is inactive, or immobile (p. 97). For brevity, I will call this the "agent is inactive" interpretation. But Norcross rejects the resulting version of HARM because of the following case.

> *Button pusher 2*: An agent, named Agent stumbles onto an experiment conducted by a twisted scientist, named Scientist. He is seated at a desk with one hundred buttons, numbered "0" through "99," in front of him. He tells her that the buttons control the amount of pain to be inflicted on a victim, named Victim. If no button is pressed within the next thirty seconds, Victim will suffer excruciating agony. If the button marked "99" is pressed, Victim will suffer slightly less; if "98" is pressed, Victim will suffer slightly less, and so on, down to "0," which will inflict no suffering on Victim. He was, he explains, about to sit and watch as Victim suffered the maximum amount. However, to honor her arrival, he turns control of the buttons over to Agent. She is free to press any button she wishes, or to press none at all. Agent pushes "99", inflicting almost maximal suffering on Victim. (p. 97)

According to HARM with the "agent is inactive" interpretation, Agent has benefited Victim, since Victim is better off, as a result of Agent's pressing "99," than she would have been if Agent had done nothing. But Norcross rejects this view, holding that "surely her act doesn't benefit Victim. It led to excruciating agony for him, when he needn't have suffered at all. She could have pressed '0' instead. If any act harms, it seems clear that this one does" (p. 98). I will dispute this verdict, at length, below.

The second interpretation of the non-performance clauses that Norcross considers is that the agent was absent, or not even "on the scene" (p. 98). I will call this the "agent is absent" interpretation. Norcross points out that the resulting version of HARM follows the previous version in implying that Agent has benefitted Victim, since Victim is better off, as a result of Agent's pressing "99," than she would have been if Agent had been absent (p. 98). So, on Norcross's view, this version of HARM has the same fatal flaw as the previous one. (I will dispute this verdict as well.)

The third interpretation of the non-performance clauses that Norcross considers is the "closest possible world" interpretation: "Instead of comparing the world in which the act occurs with a world in which the agent is either immobile or absent from the

scene, we compare it with a world that is as much like it as possible, consistent with the act not occurring" (p. 99). This interpretation will often give different results from both the "agent is inactive" interpretation and the "agent is absent" interpretation, because in many situations, the closest possible world in which the agent does not perform the act in question is one in which they perform some *other* act, rather than being inactive or absent (p. 99).

As with the previous two versions of HARM, Norcross rejects this one by way of counterexample – in this case, quite convincingly, in my view. For an especially vivid case, suppose a person irritates the famously intemperate basketball coach Bobby Knight, and he chokes her violently. And suppose, in addition, that in the closest possible world in which he does not do that, he assaults her even more violently – the only reason he does not actually do that worse act (in the hypothetical) is that he has been taking anger-management classes (pp. 99–100). According to HARM with the "closest possible world" interpretation, Knight actually benefits the person, because the person is better off, as a consequence of his choking her, than she would have been in the closest possible world in which he does not perform that specific act. But it is clearly preposterous to judge that Knight benefited her just because his personality was such that, if he had not done that, he would have done something worse (p. 100).

Moreover, generalizing from this example, we can see that the more malevolent a person is, the more likely HARM with the "closest possible world" interpretation is to judge any given maleficent act by them as a benefiting, because their personality makes it likely that in the closest possible world in which they do not do *that* maleficent act, they do something even worse. Conversely, the more benevolent a person is, the more likely this version of HARM is to judge any given beneficent act by them as a *harming*, because their personality makes it likely that in the closest possible world in which they do not do *that* beneficent act, they do something even better (p. 100). So, this version of HARM is deeply flawed.

Norcross's critique of HARM features many other valuable discussions, including several thought experiments (pp. 90–96) that elaborate ingeniously on Derek Parfit's discussions of harm in *Reasons and Persons* (Parfit 1984), and some remarks that, Norcross notes (p. 87, n. 5), build on a reply that he gave to a paper on Parfit that I presented in 1999 (subsequently published as Eggleston 2003). Although Norcross's further thoughts on Parfit's discussions illuminate those topics deeply, here I focus on his critique of HARM and his proposed alternative.

## 2.3. Norcross's contextualist account of harm

Norcross's critique of HARM, although forceful and multifaceted, is open to objection, as I argue below. First, though, let me provide a brief overview of the alternative account of harm that he presents. He characterizes this alternative as "contextualist … similar in form to some recent contextualist approaches to the epistemological notions of knowledge and justification" (p. 111). This approach leads to the following principle, which Norcross defends as superior to HARM:

> **H-con:** An act A harms a person P iff it results in P being worse off than s/he would have been had the appropriate alternative been performed. (p. 117)

This, of course, raises the question of how to identify the appropriate alternative. Norcross holds that this is determined, at least in part, by "[f]eatures of the conversational context

in which a particular action is being assessed" (p. 117). This is because "[f]eatures of the context of utterance help to determine the referents of the terms" (p. 119).

For an illustration of the role of conversational context, consider the following example. Norcross's father writes a will specifying that Norcross will receive half of his estate. If his father had not written a will, he would have died intestate, and Norcross would have received the entire estate. Also, however, Norcross has exhibited a "lack of filial piety" and does not deserve any of the estate (p. 102). About this case, Norcross writes the following:

> Does my father's act of will-writing harm me or benefit me? Imagine a conversation focused on my previous plans to invest the whole estate, based on my expectation that I would receive the whole estate. In such a context, the appropriate alternative, with which to compare the actual act of leaving me half the estate, may well be one in which I receive all the estate. In which case, my father's act resulted in me being worse off than I would have been, had he performed the appropriate alternative. Thus, the claim that my father harmed me by writing the will may express a true proposition. Imagine, though, a different, but equally natural, conversation focusing on my lack of filial piety and the fact that I clearly deserve none of the estate. In this context, the appropriate alternative may well be one in which he writes me out of his will altogether. In which case, my father's act resulted in me being better off than I would have been, had he performed the appropriate alternative. Thus, in this context, the claim that my father benefitted me by writing the will may express a true proposition. (p. 118)

Exactly how conversational context determines the appropriate alternative "may be … complicated," Norcross writes (p. 119), but he notes that one major factor is "salience" – "roughly, the degree to which the participants in a conversational context consciously focus on an alternative" (p. 120).

What are the implications of this view for statements of the form "act A harms person X"? Well, it is fine for people to utter statements that have that form. But such statements are always shorthand for statements of the form "act A results in person X being worse off than X would have been had the appropriate alternative been performed." And because the appropriate alternative is always determined by conversational context, there are no fundamental facts of the form "act A harms person X."

## 3. Some concerns about Norcross's arguments

Norcross's critique of the standard account and his contextualist account work in tandem: the supposed shortcomings of the standard account open the door to a novel account, and the supposed plausibility of the novel account releases us from needing to accept the standard account with all its supposed shortcomings. Accordingly, my defense of the standard account will include some concerns about Norcross's contextualist account. I will begin with those.

### 3.1. Some concerns about Norcross's contextualism

#### 3.1.1. Harm ascriptions and reports of harm ascriptions

We have seen that Norcross does not seek to eliminate the word 'harm' from our everyday moral discourse. In this way, the implications of his contextualist account are not as radical as they could be. Nevertheless, some of the implications are quite strange. For

example, imagine a person who engaged in two conversations, on consecutive days, about whether a particular prior act was a harming or a benefiting (or neither). If Norcross's contextualist account were correct, then it could have been correct for the person to say the following in the second conversation:

> In a conversation yesterday, in full knowledge of the facts of the act and its consequences, I called the act a harming. That claim expressed a true proposition. But the act was a not a harming.

The way Norcross's account would license this statement is as follows. In the first conversation, the appropriate alternative might have been a better act, so it would have been correct (according to Norcross's account) for the speaker to have called the act a harming. And it would then have been correct for the speaker to report that fact subsequently, such as in the second conversation. But in that second conversation, the appropriate alternative might have been an equivalent or worse act, so it would have been correct (according to Norcross's account) for the speaker to have denied that the act was a harming. And it would have been correct (on that contextualist account) for the speaker to have denied that it was a harming, period – not just to have denied that it was worse relative to some alternative.

In response to this sort of example, a defender of Norcross's account might say that if statements such as the one given above are conversationally infelicitous, it is because the speaker fails to do their interlocutors the courtesy of pointing out the difference between the two conversational contexts. That is a fair point. However, my point is not that such statements are conversationally infelicitous (though I think they are). My point is that such statements strongly appear to express false propositions, in the same way as statements such as "Yesterday I said the act was done in Colorado, and that claim expressed a true proposition, but the act was not done in Colorado." Yet Norcross's account implies that statements such as the one given above can express true propositions.

### 3.1.2. The information needed for making correct harm ascriptions
Second, consider the information that a speaker would need to know in order to justifiably call an act a harming or a benefiting. Presumably the speaker would need to know things such as what the act consisted of, who was affected, how they were affected, how they would have fared (and what else would have occurred) if the act had not been performed, and so on. Such information might be hard to ascertain, but it seems plausible to think that if a speaker were to somehow come into possession of *all* of it, they would certainly be well-positioned to discern whether the act was a harming or a benefiting and characterize it accordingly.

However, on Norcross's view, the speaker needs more information. In particular, the speaker needs to know what the "appropriate alternative" is. And that is not necessarily a straightforward matter – recall Norcross's statement that it can be "complicated." To be clear, the problem is not that being in a position to accurately call an act a harming or a benefiting might be more epistemically demanding than we might have assumed – we should be open to accounts of harm that imply that. The problem is that the additional information required has nothing to do with what the act consisted of, its effects, and so on, because it is about the conversation that the speaker is participating in.

The oddity of this implication of Norcross's account can be made more vivid with a variant of the scenario we considered a few paragraphs ago. Imagine a person who

engaged in two conversations, on consecutive days, about whether a particular prior act was a harming or a benefiting (or neither). If Norcross's contextualist account were correct, then it could have been correct for another person to say the following in the second conversation:

> Today you called the act a harming, and your claim expressed a true proposition. I know you said the same thing yesterday, but I don't know whether your claim expressed a true proposition then, because I don't know what proposition your claim expressed then.

This would be a very odd thing for a person to say – why should there be any doubt about what proposition the first person's claim expressed, the previous day? But on Norcross's account, the second person would indeed be mired in such uncertainty unless they had sufficiently detailed information about the conversation that occurred the previous day – even if the second person knew everything about what the act consisted of, its effects, and so on.

### 3.1.3. Ordinary speakers' semantic competence

My final concern is more general and fundamental than the previous two. It is that Norcross's view implies that people generally misunderstand the semantics of the word 'harm'. People generally think that the proper application of the word depends on what the act being considered consisted of, its effects, and so on – not the conversational context in which it is used. For example, people generally think that if they were to call an act a harming in some conversation and managed (in so doing) to say something true, then they would be speaking truly in calling the same act a harming in any other conversation. But Norcross's view implies that people are mistaken in their thinking on this very basic point.

Norcross directly addresses the question of whether it is plausible that people are generally mistaken on this kind of point. He does so in the context of considering whether his critiques of such notions as rightness, goodness (as applied to actions), and harm should lead to "a form of eliminativism, combined with an error theory regarding our common usage of these terms" (p. 109). Although he ends up eschewing those options in favor of his contextualism (p. 110), he does not fault the error theory for implying that "we have *all* been mistaken all this time" (p. 109). On the contrary, he writes that "I don't find this possibility particularly implausible. Similar things may well be true for certain areas of theological or scientific discourse." For example, the fact that many people believe in the existence of a god is little reason to hold that belief (pp. 109–10). Similarly, the famous example of phlogiston illustrates that even when the best scientists believe something to exist, it might not (p. 110). In Norcross's opinion, these examples are closely analogous to the issues at hand:

> If we discover that certain commonly accepted moral categories, such as right, wrong, permissible, harm, and the like, don't, in fact, pick out any fundamental moral facts, that would really be no more surprising than discovering that the gods, or fundamental physical particles (or physical properties), whose existence seems to be assumed by the discourse of theologians or scientists don't really exist. (p. 110)

However, there is a significant difference between people being mistaken about what entities exist and people being mistaken about the basic functioning of accepted moral categories such as right and harm.

On Norcross's view, "claims about harm (and benefit) have an indexical element" (p. 119). Is it plausible that people are generally mistaken about this aspect of those terms? Although perhaps most people do not know the word 'indexical', most people do have a firm grasp of the concept in practice. For example, suppose a person were to hear two people say the following:

(1) I knew Norcross's father for 5 years.
(2) I knew Norcross's father for 10 years.

Any competent speaker of English would also understand that the two speakers were not making inconsistent claims, because the word 'I' was being used to refer to different people. Similarly, suppose a person were to hear two people say the following:

(3) The weather here is great.
(4) The weather here is awful.

Again, any competent speaker of English would understand that they could interpret the speakers as either disagreeing with each other, or not disagreeing because they were referring to different places. (And perhaps they would have facts at hand that would immediately point to one interpretation, or the other.) Given these examples, we should be suspicious of arguments requiring us to attribute a significant degree of ignorance to competent speakers of English regarding the indexical elements of certain terms.

Norcross's examples of gods and phlogiston are relevant to the general phenomenon of widespread error. But some kinds of widespread error are more plausible to assert than others, as I will discuss more fully in the conclusion. The claim that words such as 'harm' and 'benefit' have an indexical element would be more plausible if one could point to more specifically relevant examples, such as terms whose indexical elements used to be obscure to people for some reason, but that are now generally acknowledged. It is unclear whether such examples can be found.

As I mentioned above, Norcross characterizes his contextualist account as similar to some recent contextualist views in epistemology. In this connection it is worth noting that even proponents of such views tend to admit that they are views that "one needs to be argued into: it takes *work* to come to think" that the truth of '*S* knows that *p*' might depend on factors other than the content of *p* and the psychological states and other characteristics of *S* (Rysiew 2023). Similarly, the radical implications of Norcross's account suggest that it, too, is something that we need to be "argued into." As I mentioned above, part of this argumentative work consists of a critique of the standard account. Does it have such grievous shortcomings that we should even be interested in an alternative? Let us consider this question next.

### 3.2. Some concerns about Norcross's critique of the standard account

As discussed above, the standard account of harm is based on the following principle:

> HARM: An act A harms a person P iff P is worse off, as a consequence of A, than she would have been if A hadn't been performed. An act A benefits a person P iff P is better off, as a consequence of A, than she would have been if A hadn't been performed. (p. 83, quoted above)

We also saw above that Norcross's critique of this principle is organized around his survey of three possible interpretations of the clauses referring to the nonperformance of the act: the "agent is inactive" interpretation, the "agent is absent" interpretation, and the "closest possible world" interpretation. I grant that in putting forward these three possibilities, Norcross has charitably encompassed all of the things that a proponent of HARM might plausibly have in mind. In addition, I agree that the version of HARM resulting from the third interpretation is implausible, for the reason illustrated by Norcross's example of Bobby Knight: this version of HARM makes the question of whether a particular act was a harming depend, to an implausible degree, on the character of the agent and what the agent's "next most likely act" would have been.

However, I deny that Norcross succeeds in showing the implausibility of the first two versions of HARM – the ones resulting from the first two interpretations of the nonperformance clause. As we saw, in both cases his argument rests on classifying the act in *Button pusher 2* as a harming. Recall that in this case, Scientist has set up a machine to inflict pain on Victim, but gives Agent the opportunity to lessen or eliminate the pain, as follows: If Agent pushes button "99," the pain will be slightly less than if agent pushes no button; if Agent pushes "98," the pain will be slightly less than with "99," and so on down to "0," which will result in no pain at all. Agent pushes "99."

We saw that according to each of the first two versions of HARM, Agent benefits Victim, since Victim is better off, as a result of Agent's pressing "99," than she would have been if Agent had been inactive or absent. But, as we also saw, Norcross rejects this conclusion. I believe Norcross's claims are vulnerable to several objections. (In what follows, I will defend the first two versions of HARM equally, and not argue for one above the other.)

### 3.2.1. Norcross's commentary on Button pusher 2, part 1

Recall Norcross's assessment of Agent's act: he writes that "surely her act doesn't benefit Victim. It led to excruciating agony for him, when he needn't have suffered at all. She could have pressed '0' instead. If any act harms, it seems clear that this one does" (p. 98, quoted above).

In this passage there are several remarks that might be read as supporting the verdict that the act is a harming. The first is "It led to excruciating agony for him." What support for the harming verdict does this fact provide? Well, it might be a prima facie indicator of the act being a harming, since acts that non-accidentally cause pain are almost always harmings. Relatedly, one might see the act as a harming because it is indeed a harming in a certain respect, and one might forget the distinction between that notion and the notion of harming all things considered. But as we saw in Norcross's case of *Doctor*, pain-causing acts can be all-things-considered benefitings in unusual circumstances, such as ones in which some pain or other is already on track to happen anyway. Thus, the *Doctor* case tells us that in a case such as *Button pusher 2*, we cannot take the fact that Agent's act caused pain for Victim as indicating that it harms Victim.

The second stretch of text that might be read as supporting the harming verdict is "he needn't have suffered at all. She could have pressed '0' instead." To be sure, that would have been optimal for Victim, and thus more beneficial than pressing any other button, including "99." But that does not mean that pressing any other button was not beneficial, too, albeit to a lesser degree.

The third and final remark that might be read as supporting the claim that the act is a harming is "If any act harms, it seems clear that this one does." It is natural to think that when an agent has many different possible acts (such as one hundred of them, as in

this case), some of the acts will be harmings and some will be benefitings. So, if we know which act is the worst along the benefit–harm continuum, then it seems reasonable to conclude that it is a harming (maybe the only harming, maybe not). Like the observation about non-accidentally caused pain almost always being an indicator of harming, the principle just mentioned might be a generally reliable heuristic for correctly categorizing acts as harmings. But we know that *Button pusher 2* has features that make heuristics unreliable. Returning to Norcross's concluding remark "If any act harms, it seems clear that this one does," we can say that it might be true, but only because the antecedent is false, because in this case, there are no harmful acts in the set of options.

It seems, then, that the passage quoted above provides little support for Norcross's verdict that that Agent's act is a harming. To be sure, Agent's act should elicit a fusillade of criticism from any reasonable observer: it was the worst thing Agent could do, aside from doing nothing; pressing "99" was not any easier, physically, than pressing "0" would have been; if it would have been psychologically more difficult for Agent to press "0," there is something very wrong with Agent's motivations; only a monster would deliberately choose "99"; etc. But as Norcross himself notes (pp. 84, 95, and 127), we need to bear in mind that judging an act to be a benefiting need not redound to the agent's credit. Thus, all of the criticisms just mentioned above can be said *alongside* – not *instead of* – the verdict that pressing "99" was, precisely speaking, a benefiting.

### 3.2.2. Norcross's commentary on Button pusher 2, part 2

In a later paragraph, Norcross writes that *Button pusher 2* points to a deeper problem with the first two versions of HARM: "The problem is not just that inactivity [and presumably absence] gives unacceptable results in particular cases, but rather that the comparisons it invites do not seem relevant to whether an act harms or benefits" (p. 98). Continuing in this vein, Norcross asks why it should "matter" in a case like *Button pusher 2* that Victim "would have suffered even more if [Agent] had been immobile or absent" (p. 98), and says the verdict of harming or benefiting in such cases "doesn't seem to depend" on the counterfactuals involved in the first two versions of HARM (p. 99). He concludes, "These counterfactuals, then, don't seem relevant to the question of whether an act harms or benefits" (p. 99).

These claims, if true, would indeed be devastating to the first two versions of HARM. However, these claims essentially restate the overall dispute at hand, and it is open to a defender of HARM to simply deny that the claims are true. It would take us too far afield to explore the abstract, general reasons one might give for and against the relevance of the counterfactuals Norcross mentions above. Instead, let us keep things more specific and intuitively provocative by using variations on *Button pusher 2* to further evaluate Norcross's assessment of that case, and to further evaluate Norcross's contextualism vis-à-vis HARM.

### 3.2.3. The case of Angela and Vincent (Button pusher 2, reversed)

Let us further evaluate Norcross's assessment of *Button pusher 2* by considering a variant of that case. The variant will be structurally identical to *Button pusher 2*, but with all the valences reversed, and we will adapt Norcross's reasoning accordingly.

*Angela and Vincent*: Angela stumbles onto an experiment conducted by Scientist. He seats her at a desk with buttons numbered "0" through "99." If no button is

pressed within the next thirty seconds, Vincent will enjoy extreme pleasure. If the button marked "99" is pressed, Vincent will have slightly less pleasure; if "98" is pressed, Vincent will have slightly less pleasure than with "99"; and so on, down to "0," which will result in no pleasure at all. Scientist was about to sit and watch as Vincent enjoyed the maximum amount of pleasure. However, to honor Angela's arrival, Scientist gives Angela the option of pressing any button, or none at all. Angela pushes "99", providing Vincent with almost maximal pleasure.

Now, what assessment of this case would parallel Norcross's assessment of Button pusher 2? It would seem to be the following:

[S]urely her act doesn't harm Vincent. It led to extreme pleasure for him, when he needn't have had any at all. She could have pressed "0" instead. If any act benefits, it seems clear that this one does.

I submit that this assessment is not compelling. First, although Angela's act led to extreme pleasure, doing nothing would have been even better. Second, the fact that she could have pressed "0" and thereby done much more harm does not mean she did not do any harm by pressing "99." Finally, it is very plausible to think that none of her options would have been a benefiting – from which it follows that this option, by being the best of the bunch, is still not a benefiting. This case, then, gives us further reason to doubt Norcross's assessment of *Button pusher 2*.

In contrast, this case exposes nothing questionable about HARM on either the "agent is inactive" interpretation or the "agent is absent" interpretation. Both of these versions of HARM imply that Angela harms Vincent. This is straightforwardly intuitive: Angela took away a bit of the pleasure that Vincent was going to enjoy.

### 3.2.4. The case of 111 Buttons

So far in this section, I have been rebutting Norcross's assessment of *Button pusher 2*, in order to defend the first two versions of HARM against his critique of them. Now, I would like to use one final variation on *Button pusher 2* to cast further doubt on the plausibility of Norcross's contextualism vis-à-vis HARM.

*111 buttons*: Everything is the same as in *Button pusher 2*, except that there are 11 additional buttons, numbered "100" through "110." If the button marked "100" is pressed, Victim will suffer the same agony as if no button is pressed; if "101" is pressed, Victim will suffer slightly more; if "102" is pressed, Victim will suffer slightly more than with "101"; and so on, up through "110," which is the most painful possible outcome in this scenario. As in *Button pusher 2*, Agent pushes "99."

Now, what would Norcross's contextualism say about whether Agent's pushing "99" was a harming? Neither answer – a harming or not a harming – works out well for the view. If the view holds that the act is a harming, that verdict at least has the virtue of maintaining consistency with its verdict of harming in regard to *Button pusher 2*. But we saw above that that verdict is counterintuitive, and the same verdict is arguably even more counterintuitive now that buttons 100–110 are available. After all, would not

pushing buttons 101–110 be harmings – acts that make things worse for Victim – and is not pushing button 99 *not* a harming because it makes things better for Victim?

On the other hand, suppose Norcross's contextualism holds that pushing "99" is a benefiting in this case, perhaps on the grounds that the addition of buttons 100–110 changes the conversational context of assessment. The verdict of benefiting is itself perfectly intuitive, but the juxtaposition of *Button pusher 2* and *111 Buttons* raises questions about Norcross's contextualism being over-sensitive to additions to agents' option sets, if indeed such additions are deemed to change the conversational context of assessment. If adding worse buttons can flip the verdict for pushing "99" from harming to benefiting, how would that verdict be affected if we kept all 111 buttons but added some buttons that would lead to pleasure rather than pain? Would that flip the verdict back to harming? I think we intuitively expect the categories of harming and benefiting to exhibit a certain kind of consistency or stability across cases, and that expectation seems to cast doubt on any view that calls pushing "99" a harming in *Button pusher 2* but calls it a benefiting in *111 Buttons*.

In contrast, the first two versions of HARM handle the case of *111 Buttons* simply and intuitively, and consistently with *Button pusher 2*. In *111 Buttons*, pushing "99" remains a benefiting, albeit an intentionally minimal one reflecting poorly on Agent's character.

## 4. Conclusion: some intuitions matter more

Let us review. Norcross proposes that when someone characterizes an act as a harming, they should be understood as saying that the act was worse than the appropriate alternative, as determined by conversational context. Recalling Norcross's primary example from above, if a person says that Norcross's father's act of will-writing did not harm Norcross (despite the fact that it lessened his inheritance), the utterance would be true if the conversational context established that the appropriate alternative was his inheriting nothing at all. On this contextualist view, there are no fundamental facts of the form "act A harms person X." Such statements do not need to be purged from ordinary conversation, but they do need to be reinterpreted as making comparative claims, as in the example just given.

To motivate this revisionary view, Norcross criticizes the standard account, stated by HARM. He plausibly claims that this principle can be understood in any of three ways, and he argues convincingly that the "closest possible world" version of that principle is implausible. To argue against the other two versions of that principle – the "agent is inactive" version and the "agent is absent" version – Norcross relies heavily on *Button pusher 2*. He argues that Agent's pushing "99" is not a benefiting, as the first two versions of HARM imply. I rebutted this assessment at length, above.

The contest between HARM and Norcross's view mostly boils down to a clash of intuitions. Norcross claims that HARM's verdict in *Button pusher 2* is counterintuitive, while I claim that many implications of his contextualism are counterintuitive. Probably the most notable such implications of his view are not related to any specific example, but are simply the claim that the words 'harm' and 'benefit' have an indexical element and the ensuing blanket denial of the existence of fundamental facts of the form "act A harms person X." These implications are radically at odds with how we ordinarily think about harm, regardless of how one assesses specific situations such as *Button pusher 2*.

The extensive use of objections based on counterintuitive implications, by both Norcross and myself, warrants some concluding comments. As I indicated earlier,

Norcross is aware of the counterintuitive implications of his view and holds that we should not be constrained by intuition when ascertaining the merits of philosophical views. He is admirably principled on his point, even holding that when forms of utilitarianism and other forms of maximizing consequentialism that he rejects are criticized for having counterintuitive implications, their proponents are justified in rejecting those objections (pp. 16–20, 37–40, and 141–42). His steadfast insistence that his own view should be accepted, despite its counterintuitive implications, reminds one of the definition of the term 'outsmart' in *The Philosophical Lexicon*:

> To embrace the conclusion of one's *reductio ad absurdum* argument. "They thought they had me, but I outsmarted them. I agreed that it *was* sometimes just to hang an innocent man." (Dennett 1987, p. 15)

It is no surprise that the eponym of this term, J. J. C. Smart, was a proponent of act utilitarianism, just as Norcross is (though they diverge, of course, on the choice between maximizing and scalar forms of the view). Because act utilitarianism is probably the most common target of such *reductios*, its defenders have more occasion than most moral theorists to find that "outsmarting" their opponents, despite its obvious costs in terms of intuitive plausibility, might be their best option.

However, not all intuitions are on a par, and equally eligible for being overruled by philosophical theorizing. It is one thing to reject a widely held judgment about a specific situation, as in the example given in the definition of the term 'outsmart', above. Another example, more germane to our present topic of harm and benefit, would be the intuition, held by most people, that one is harmed by infidelity by one's spouse, even if one never finds out about it and there is no other effect on one's subjective experience. Again, it is one thing to challenge that first-order substantive intuition of harmfulness, as hedonists about well-being do. It is quite another thing to suggest that we radically revise our understanding of the meaning of such concepts as harming and benefiting. Yet this is exactly what is asked of us by Norcross's contextualism, with its assertion of an indexical element for those concepts and its denial of the existence of fundamental facts of the form "act A harms person X."

The distinction I am drawing, between challenging substantive intuitions about specific cases and challenging our understanding of the meaning of core concepts, loosely parallels a distinction that R. M. Hare makes. As yet another act utilitarian, Hare was forced to reckon with the counterintuitive implications of the view, and he was generally unstinting in his defense of them. For example, his defense of utilitarianism's refusal to categorically condemn slavery (Hare 1989, pp. 158–62) might well be cited as another example of "outsmarting." In support of this style of argument, Hare held that "Common moral opinions have in themselves no probative force whatever in moral philosophy" (Hare 1972, p. 122). But he was much more deferential toward commonly held opinions about the meanings of moral words. If a theory is revisionary in this realm, it is on shakier ground (Hare 1972, pp. 121–22). Indeed Norcross himself implies some sympathy for this view when he rejects a certain theory of rightness because it requires us to assume its opponents "are simply confused about the meanings of words," which he writes "is highly implausible" (p. 35).

Norcross's contextualism might or might not run afoul of Hare's warning – the exact contours of the realm that Hare had in mind are not entirely clear. But Hare's view should at least make us more attuned to the different kinds of opinions that ethical theories might ask us to revise. And surely a high burden of proof must be met by any

theory that asks us to radically revise our common understanding of core features of concepts such as harming and benefiting. The considerations presented in this article suggest that Norcross's view does not meet that burden, and thus that moral theorists need not be deterred by Norcross's arguments from continuing to rely on the standard account of harm, and using it to make non-comparative claims of form "act A harms person X."

## References

**Bradley, Ben**. 2012. "Doing Away with Harm," *Philosophy and Phenomenological Research* **85**, pp. 390–412.

**Carlson, Erik, Jens Johansson, and Olle Risberg**. 2022. "Causal Accounts of Harming," *Pacific Philosophical Quarterly* **103**, pp. 420–45.

**Dennett, Daniel**. 1987. *The Philosophical Lexicon*, 8th ed. Newark, Delaware: American Philosophical Association. URL = <https://dl.tufts.edu/concern/pdfs/jw827p68f>, last accessed March 28, 2024.

**Eggleston, Ben**. 2003. "Does Participation Matter? An Inconsistency in Parfit's Moral Mathematics," *Utilitas* **15**, pp. 92–105.

**Gardner, Molly**. 2021. "What Is Harming?" in *Principles and Persons: The Legacy of Derek Parfit*, ed. by Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan (Oxford: Oxford University Press), pp. 381–95.

**Hare, R. M.** 1972. "The Argument from Received Opinion," in R. M. Hare, *Essays on Philosophical Method* (Berkeley: University of California Press), pp. 117–35.

**Hare, R. M.** 1989. "What Is Wrong with Slavery," in R. M. Hare, *Essays on Political Morality* (Oxford: Oxford University Press) pp. 148–66.

**Mill, John Stuart**. 1859. "On Liberty," in *The Collected Works of John Stuart Mill*, ed. by J. M. Robson, 33 vols. (Toronto: University of Toronto Press, 1963–91), vol. **XVIII**, pp. 213–310.

**Norcross, Alastair**. 2020. *Morality by Degrees: Reasons without Demands* (Oxford: Oxford University Press).

**Parfit, Derek**. 1984. *Reasons and Persons* (Oxford: Oxford University Press).

**Rysiew, Patrick**. 2023. "Epistemic Contextualism," in *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), ed. by Edward N. Zalta and Uri Nodelman, URL = <https://plato.stanford.edu/archives/win2023/entries/contextualism-epistemology/>, last accessed March 28, 2024.