

## A Defense of the Harm-Based Solution to the Non-Identity Problem

---

### Introduction

What has come to be known as the *non-identity problem* raises some notoriously difficult questions about the relationship between harm and well-being. While the literature now contains an impressive array of strategies for solving the problem,<sup>1</sup> I will focus in this paper upon one particular strategy, which I call the *harm-based approach*. Roughly put, the harm-based approach holds that certain procreative actions can harm the resultant individuals, even if the lives of those individuals are well worth living. One way to try to justify this approach is to endorse the *non-comparative account of harming*, which holds that an event harms an individual just in case it causes her to be in a bad state, such that the state's badness does not derive from a comparison between that state and some alternative state that the individual would or could have been in. Many philosophers argue that this conception of harming is inadequate,<sup>2</sup> and one might be tempted by such critiques to draw that conclusion that any harm-based approach to the non-identity problem will fail.

In this paper, I argue that the non-comparative account of harming is, indeed, inadequate, but this need not impugn all harm-based approaches to the non-identity problem. I will begin by

<sup>1</sup> For a helpful overview of proposed solutions, see Roberts (2013).

<sup>2</sup> Philosophers who argue that the non-comparative account is an inadequate account of harming include Matthew Hanser (2008), Judith Thomson (2010), and Ben Bradley (2012). Also, for reasons I don't discuss in this paper, some philosophers argue that the concept of non-comparative harming will not actually solve the non-identity problem, as it is advertised to do. These critics include Fiona Woollard (2013), Derek Parfit (1987), and David Boonin (forthcoming). For discussion of this latter set of objections to non-comparative harming, see my unpublished manuscript.

setting forth the non-identity problem in more detail. I will then explain why the non-comparative account of harming, though initially an attractive response to the non-identity problem, is ultimately doomed to failure. Finally, I will defend what I call the *existence account of harming*. I will conclude that this account can vindicate the harm-based approach to the non-identity problem without relying on the concept of non-comparative harming.

## **1. The Non-Identity Problem and the Non-Comparative Account of Harming**

Consider the following two cases:

*Case 1.* Alice is a troubled woman who wants sympathy. She believes that if she has a child with poor health, she will receive sympathy from her friends and family. During her pregnancy, she takes a drug that causes her child, whom she names Alex, to develop poor health. Despite his health, Alex has a life worth living. He would have had a higher level of well-being if Alice had not taken the drug.

*Case 2.* Barbara is a troubled woman who wants sympathy. She believes that if she has a child with poor health, she will receive sympathy from her friends and family. Barbara uses *in vitro* fertilization and screens the embryos for a particular gene that causes poor health. When she finds an embryo with that gene, she implants it and discards the rest. The selected embryo becomes a child named Billy, who develops poor health. Having poor health causes Billy to experience exactly the same hardships, pain, and suffering that having poor health causes Alex to experience. However, like Alex, Billy has a life worth living.

Intuitively, both Alice's action and Barbra's action are objectionable. The objection to Alice's action is that she has clearly harmed her child; she has brought it about that her child experiences poor health. Since Barbara has also brought it about that her child experiences health that is equally poor, we might be tempted to think that the objection to Barbara's action is also grounded in harm.

Nevertheless, there is a difference between Case 1 and Case 2. The difference is that, although Alex would still have existed, had his poor health *not* been induced, Billy is *non-*

*identical* to anyone who would have existed, had his poor health not been selected for. After all, if Barbara had not selected for poor health, then either she would not have had a child, or else she would have implanted a different embryo, and a different person would have come into existence.

Many philosophers argue that this metaphysical difference makes a moral difference. According to the *counterfactually worse-off condition* on harming, an action harms someone only if it makes her worse off in some respect than she would have been, had the action not been performed. Alice's action satisfies this condition: Alex is worse off in many respects than he would have been, had Alice not induced his poor health. However, Barbara's action does not satisfy this condition. Billy's life is worth living, and plausibly, having a life worth living is not worse for Billy in any respect than not existing; therefore, Billy is no worse off in any respect than he would have been, had Barbara not selected for poor health. Thus, the counterfactually worse-off condition implies that Barbara does not harm Billy.

But if Barbara's action does not harm Billy, then we seem to be at a loss to justify the intuition that, in much the same way that Alice's action is objectionable, Barbara's action is also objectionable. Case 2 is thus an example of a *non-identity case*: it is a case in which we find it difficult to explain why an individual appears to be wronged by an action that is the condition of his or her own worthwhile existence. The problem of either accounting for this appearance that the individual was wronged or explaining it away is the *non-identity problem*.<sup>3</sup>

Here I want to clarify how I am using the term “wronged.” Some people hold that when an action “wrongs” an individual, the action is all-things-considered wrong. However, many

---

<sup>3</sup> The term, “non-identity problem” was coined by Derek Parfit (1987), but the fundamental tensions that constitute the problem were identified, somewhat independently, by Parfit (1976), Robert Adams (1979), Thomas Schwartz (1978), and Gregory Kavka (1982).

philosophers who have written about the non-identity problem do not hold that in *every* non-identity case, the procreative action is all-things-considered wrong.<sup>4</sup> Therefore, the sense of “wronged” to be explained in a non-identity case is weaker. What must be explained, more precisely, is the intuition that there is a *moral reason* against the procreative action, where that reason has something to do with the well-being of the individual whose existence is conditional on the action. I will assume that if we can account for the existence of and the strength of this reason—or if we can successfully explain away the intuition that there *is* such a reason—we will have solved the non-identity problem.

The *harm-based approach* to the non-identity problem is any approach that attempts to solve the problem by rejecting the counterfactually worse-off condition on harming. Those who opt for the harm-based approach hold that an action can harm an individual even if it fails to make her worse-off in any respect than she would have been, had the action not been performed.

It is tempting to think that, in place of the counterfactually worse-off condition, the proponent of the harm-based approach ought to accept the *non-comparative account of harm*, according to which harm is a state of affairs that is bad for an individual, such that the badness of the state does not derive from its being worse for the individual than some other state that the individual would or could have been in. It is also independently plausible that harming an

---

<sup>4</sup> There is some disagreement on this point in the literature. For example, Harman (2009) describes non-identity cases differently than I do, and she contends that the intuition about such cases is that the action is all-things-considered wrong. She writes, “A ‘non-identity case’ is a case in which an action affects whether some people exist, and it appears (either to everyone, or to some people) that the action is wrong in virtue of harming these people, although they have lives that are worth living” (2009, p. 140). On the other hand, Parfit provides two different definitions of “non-identity problem” that are somewhat imprecise, but which support my claim that the intuition to be explained is that there is a *reason* against the action, not that the action is always all-things-considered wrong. He first claims that the non-identity problem is the problem of determining what is the “objection” to a 14-year-old girl’s decision to have a baby (p. 359). He later invokes a case he calls “Depletion” and states that the non-identity problem is the problem of answering the question, “What is the moral reason not to choose Depletion?” (p. 363).

individual is equivalent to causing that individual to suffer harm. The combination of these two claims yields the following:

### **The Non-Comparative Account of Harming<sup>5</sup>**

*Harming* (def.): An event E harms an individual S just in case E causes a state of affairs that is a harm for S.

*Harm* (def.): A state of affairs T is a harm for an individual S just in case T is non-comparatively bad for S.

The non-comparative account of harming seems to justify the claim that in Case 2, Barbara harms Billy. After all, Billy's having poor health is a state of affairs that is bad for Billy, and Barbara's action causes that state of affairs.

Despite the promise that it holds as a potential solution to the non-identity problem, the non-comparative account of harming faces a number of problems. First, there appear to be cases in which an action or event causes an individual to be in a state that is bad for her without harming her. Second, there appear to be cases in which an action or event harms an individual without causing her to be in a state that is bad for her. Third, it is difficult to determine just what it is that makes a state of affairs non-comparatively bad *for some individual*, especially in cases where a similar state of affairs would not be bad for another individual.

The first problem is motivated by cases like the following:

*Dim Vision.* Jones has been blind for many years as a result of retinal damage. Recently, Dr. Smith has developed a surgical operation that can repair some but not all of the damage. Dr. Smith operates on Jones and improves his vision from a state of blindness to a state in which Jones can see, but not very well: Jones now has what we will call *dim*

---

<sup>5</sup> As I have formulated it, the non-comparative account of harming combines multiple ideas, a subset of which many philosophers endorse or at least discuss, but it is not an exact statement of any one philosopher's view. Philosophers who critique but do not endorse the non-comparative account of harming include Matthew Hanser (2008), Judith Thomson (2010), and Ben Bradley (2012). A philosopher who endorses something like the non-comparative account as a complete account of harming is Seana Shiffrin (1999). Finally, Elizabeth Harman (2004, 2009) and Fiona Woollard (2012) hold that there is a *conception* of non-comparative harming that serves as a sufficient but not a necessary condition for harming.

*vision.*<sup>6</sup>

It seems that Jones's having dim vision is a state of affairs that is bad for Jones. Dr. Smith causes that state of affairs. Thus, it would seem that if the non-comparative account of harming is true, then Dr. Smith harms Jones by performing the operation. However, it is intuitively clear that Dr. Smith does not harm Jones.<sup>7</sup>

The second problem for the non-comparative account is illustrated by the following:

*Death.* Sam, a healthy and happy 25-year-old, is attempting to cross the train tracks when his foot gets stuck. A train hits him and he dies instantly.<sup>8</sup>

The event that consists of the train's hitting Sam does not appear to cause a state of affairs that is bad for Sam. That is because the state of affairs that the accident brings about is one in which Sam no longer exists, and it is difficult to see how a state of affairs can be bad for an individual who no longer exists. However, it is intuitively clear that the train accident harms Sam.

One might think that if we had a more precise account of what it is for a state to be non-comparatively bad for someone, we could explain away the apparent counterexamples and thereby vindicate the non-comparative account of harming. However, a dilemma arises when we attempt to define 'non-comparative badness' more precisely. Non-comparative badness for an individual must either be absolute or relative. But if we take non-comparative badness to be an

---

<sup>6</sup> This example is an adaptation of cases that appear in Thomson (2010) and Hanser (2009).

<sup>7</sup> Harman (2009) responds to this objection by distinguishing between a "particular bad state" and a "general bad state." She claims that causing an individual to be in a general bad state is a sufficient condition for harming him. Presumably, Harman would say that dim vision is a *particular* bad state, so it does not follow from the claim that Dr. Smith caused Jones to be in such a state that Dr. Smith harmed Jones. This response is unsatisfying. The notions of generality and particularity that Harman relies upon face a problem analogous to the generality problem for reliabilism: we need either a criterion or firm intuitions to help us distinguish general bad states from particular bad states, and we lack both.

<sup>8</sup> This case is inspired by Hanser's (2008) argument that in general, state-based accounts of harm cannot successfully explain why death is a harm. A state-based account of harm is one according to which a harm is a state of affairs; the non-comparative account of harm is one such account. As an alternative to state-based accounts, Hanser favors an event-based account of harm. I do not critique his account here, but see Thomson (2010) for some objections to Hanser's view.

*absolute* notion, we seem to face further counterexamples, and if we take badness to be a *relative* notion, we either undermine the original motivation for the non-comparative account of harming, or else we face yet more apparent counterexamples.

To conceive of a non-comparatively bad state of affairs for an individual as an *absolute* notion is to hold that a state of affairs is non-comparatively bad for an individual if and only if it would be bad for *any* individual. The problem with this account is that there appear to be cases in which an event harms someone by causing him to be in a state that is bad for him, but which would *not* be bad for others. Consider the following:

*Loss of Fortune.* Jeeves was once a world-renowned physicist with extraordinary intellectual abilities. He then had a stroke and suffered brain damage. The brain damage left him with average intellectual abilities.<sup>9</sup>

It seems that having average intellectual abilities would not be a bad state for someone who was born with them. Nevertheless, the stroke *harmed Jeeves*.

An alternative way to define what it is for a state to be non-comparatively bad for an individual is to relativize the badness to something about the individual. However, it is difficult to determine what it is about the individual that the badness should be relativized to. Although Elizabeth Harman (2009) suggests that badness should be relativized to the norm for the individual's species, Loss of Fortune illustrates one problem with that approach. Jeeves is a member of the human species, and it does not *seem* to be bad for a human to have average intellectual abilities, even though coming to have average intellectual abilities is a harm for

Jeeves.<sup>10</sup>

---

<sup>9</sup> This case is modeled upon the Bertrand Russell case in Jeff McMahan (1996) and the case of the Nobel Prize winner in Hanser (2008).

<sup>10</sup> Note, however, that Loss of Fortune is not a counterexample to Harman's view. That is because Harman takes non-comparative harming to be a sufficient but not a necessary condition for harming: "An event harms someone if it causes him pain, mental or physical discomfort, disease, deformity, disability, or death" (2009, p. 139). The problem with taking non-comparative harming to be sufficient but not necessary for harming is that it leaves us with an incomplete account. As Bradley puts it, "[Harman's account] is hardly an analysis of harm. It cries out

A second problem with relativizing badness to species norms is motivated by the following case:

*Goat.* A team of biologists have created a continuum of genetically engineered embryos, such that human embryos are at one end and goat embryos are at the other. As one moves from the human end to the goat end of the continuum, the embryos contain less and less human DNA and more and more goat DNA. The biologists are wondering whether it would be morally permissible to cause the embryos in the middle of the continuum to develop into adults.<sup>11</sup>

Imagine an embryo that is roughly in the middle of the continuum. There is either an answer to the question of what species it belongs to, or there is not. If there is no answer, then, interpreted using a species-relativized definition badness, the non-comparative account of harming cannot tell us whether letting the embryo develop would constitute harming or not; the view is therefore incomplete. If there is an answer, then there must be a distinct cut-off line that distinguishes humans with goat-like qualities from goats with human-like qualities. There will also be two possible embryos, one on either side of the line, that will be almost qualitatively identical in terms of their future phenotypes and levels of well-being. Assuming that the level of well-being that is normal for humans is higher than the level of well-being for normal goats, and assuming that the level of well-being of the two individuals in the center of the continuum will be somewhere in between these two norms, it follows from the species-relativized interpretation of the non-comparative account of harming that causing one embryo to develop would constitute harming, whereas causing the other embryo to develop would not. This result is objectionably arbitrary.<sup>12</sup>

---

for unification—what do the items on the list have in common?” (p. 399).

<sup>11</sup> This case is loosely modeled on the “everworse continuum” thought experiment in Greene and Augello (2011).

<sup>12</sup> I have adapted this objection from an argument that Greene and Augello (2011) employ against objections to selecting for disability. Greene and Augello describe the continuum that stretches from humans to goats as an “everworse continuum,” and they stipulate that as you move towards the goats, well-being gets ever worse. I think that such a stipulation is unnecessary, so I have left it out.

Could non-comparative badness be relativized to the individual's will or desires?<sup>13</sup>

Neither choice is promising. A dog with cancer has neither the will nor the desire to be cancer-free, for she does not have the concept of cancer that is a necessary condition for having such a will or a desire. Nevertheless, when she gets cancer, she is harmed.

How about interests?<sup>14</sup> If we hold that a state is non-comparatively bad for someone if and only if it is contrary to one of her interests, we are only pushing the problem back a step. For we must now decide whether to take an absolute or relativistic approach to determining *which* interests an individual has. If we attribute the same interests to all sentient individuals, then we face the kind of problem that arose for making the concept of non-comparative badness absolute: the set of universal interests will inevitably be too broad or too narrow. Either we will say that blindness is contrary to the interests of a bat, and thus, that a bat's mother harms him when she conceives him (and thereby causes him to exist in a state of blindness), or we will say that having average intellectual abilities is *not* contrary to Jeeves's interests, so the stroke that causes Jeeves to have average intellectual abilities does *not* harm him.

On the other hand, if we relativize interests, we must decide what to relativize them to. If we relativize interests to species membership, desires, or will, we will again have trouble explaining why Jeeves or the dog with cancer is harmed. Suppose we relativize interests to the level of well-being that individuals *previously* had, or to the level of well-being that individuals *would have had* if some event had not occurred: we say that Jeeves has an interest in having

extraordinary intellectual abilities because having such abilities would restore his well-being to

<sup>13</sup> Seana Shiffrin (1999) and George Pitcher (1984) appear to relativize harm to a person's will, desires, or interests. Shiffrin writes that harms are "conditions that generate a significant chasm or conflict between one's will and one's experience, one's life more broadly understood, or one's circumstances" (p. 123). Pitcher writes, "An event or state of affairs is a misfortune for someone (or harms someone) when it is contrary to one or more of his more important desires or interests" (p. 184).

<sup>14</sup> Joel Feinberg (1984) defends a highly influential account of harm that relativizes harm to an individual's interests. However, he endorses a comparative account of harm.

pre-stroke levels, or because having such abilities would boost his well-being to the level at which it *would* have been if he had not had the stroke. If we take this option, we undermine the original motivation behind the non-comparative account of harming, which was to vindicate the notion that individuals can be harmed by actions that restrict their lifetime prospects of well-being even if such individuals *never had* or *would not have had* unrestricted prospects.

Thus, the non-comparative account of harming is in trouble. The account says that harming an individual is causing a state of affairs that is non-comparatively bad for that individual. But it seems that, however we define what it is for a state to be *non-comparatively bad for* someone, either we undercut the original motivation for the non-comparative account of harming, or else we invite apparent counterexamples. These problems for the non-comparative account of harming have led some philosophers to conclude that the harm-based approach to the non-identity problem is also doomed to failure. However, as we will see in the next section, the harm-based approach need not rely upon a non-comparative account of harming.

## **2. The Existence Account of Harming**

So far, I have argued that although the non-comparative account of harming seems to offer a promising solution to the non-identity problem, it fails as a general account of harming. In this section, I will defend an alternative account of harming that not only vindicates the harm-based approach to the non-identity problem, but also gets most of the right results in the cases that stymied the non-comparative account. I will begin by stating the account in full and clarifying some of its tenets. I will then show how the account vindicates the harm-based approach to the non-identity problem. Finally, I will argue that the account gets most of the right results in the

cases that made trouble for the non-comparative account.

Here is the full account:

### **The Existence Account of Harming**

*Harming* (def).: An event, E, harms an individual, S, if and only if E causes a state of affairs that is a harm for S.

*Harm* (def).: A state of affairs, T, is a harm for an individual, S, just in case the following two conditions hold:

- (i) T falls within the scope of what can be intrinsically good or bad for S; and
- (ii) If S existed and T had not obtained, then S would be better off in some respect.

Notice, first, that I define harming as causing harm, just as the non-comparative account of harming does. This is a natural way of thinking about harming, and as I will show below, even an account that includes the counterfactually worse-off condition on harming can be formulated to capture the intuition that 'to harm' is a causal verb.

Notice, second, that clause i in the definition of harm refers to what can be intrinsically good or bad for S. This clause is formulated so as to be neutral with respect to substantive accounts of well-being. That is, regardless of whether the correct account of well-being turns out to be hedonism, desire-satisfaction theory, an objective list theory, or any other account, the existence account of harming can be conjoined with it.

Although some accounts of harming distinguish between intrinsic and extrinsic harms,<sup>15</sup> my account does not require us to countenance anything that might be called an "extrinsic harm." Suppose, for example, that hedonism is true, and that Jones steals two hundred dollars from Smith. We might think that it is not intrinsically harmful for Smith to lose two hundred dollars, but it *is* intrinsically harmful for Smith to lose some happiness. The existence account of harming can capture this intuition: by stealing from Smith, Jones causes a state in which Smith is less

---

<sup>15</sup> See Bradley (2012) and Klocksiem (2012), for example. Bradley attributes the original distinction between intrinsic and extrinsic harms to Feinberg (1984).

happy by causing the state of affairs in which Smith has less money. Thus, my account can get the result that Jones harms Smith without relying on the claim that losing money is in some way extrinsically harmful.

The second clause in my definition of harm is the most important because it does the work of vindicating the harm-based approach to the non-identity problem. To see why this is so, we can compare the existence account of harming with an account of harming that is incompatible with the harm-based approach. For example, here is one way of incorporating the counterfactually worse-off condition on harming into a complete account of harming:

### **The Counterfactually Worse-Off Account of Harming:**

*Harming* (def.): An event, E, harms an individual, S, if and only if E causes a state of affairs T, to obtain, such that T is a harm for S with respect to E.

*Harm* (def.): A state of affairs, T, is a harm for an individual, S, with respect to an event E if and only if T is worse for S in some respect than some other state, U, such that U would have obtained if E hadn't happened.

In their definitions of harm, both the existence account and the counterfactually worse-off account invoke counterfactuals, but the antecedents of those counterfactuals are different. The antecedent of the counterfactual in the counterfactually worse-off account picks out the possibility that E does not obtain. On the other hand, the antecedent of the counterfactual in the existence account of misfortune picks out the possibility that S exists and E does not obtain.

Because of this difference in how the antecedents are formulated, the existence account of harming allows for the possibility that in Case 2, Billy's having the poor health is a harm for Billy; however, the counterfactually worse-off account rules out such a possibility. To see why, suppose that the following claims can both be truthfully asserted by Billy:

(B1) If I did not have poor health, then someone else would have come into existence in

my place.

(B2) If I existed but did not have poor health, then my poor health would have been cured through advanced medicine.

B1 refers to a state of affairs in which Billy does not exist, and in which, *a fortiori*, Billy is not better off in any respect. The conjunction of B1 and the counterfactually worse-off account of harm implies that Billy's having poor health is not a harm. However, B2 refers to a state of affairs in which Billy *does* exist. Since, in that scenario, he exists *and* has better health, the following claim is also true:

(B3) If I existed but did not have poor health, then I would be better off in many respects.

B3 satisfies condition ii of my definition of harm. Assuming that i is also satisfied, Billy's having poor health is a harm. Since Barbara caused that harm, Barbara harmed Billy; by accounting for the intuition that she wronged him, this result solves the non-identity problem.

Notice that my definition of harm does not require us to deny the truth of B1. Indeed, we can suppose that B1, B2, and B3 are *all* true. However, B1 is true in virtue of one counterfactual state of affairs, and B2 and B3 are true in virtue of another; we might say that B2 and B3 refer to a possible world that is more distant than the world that B1 refers to. The mistake that proponents of the counterfactually worse-off condition on harming were making is that they were making comparisons to the *wrong possible world*. They were imagining how things would have been if the *cause* of Billy's poor health hadn't obtained. Instead, they ought to have considered how things would have been *for Billy* if *Billy's poor health* hadn't obtained.

One might object that condition ii in my definition of harm is *ad hoc*. What fuels the non-identity problem, in the first place, is the fact that Billy would not have existed if Barbara had not selected the embryo on the basis of its gene for poor health. The objection contends that on my

account, we need not consider this fact at all in assessing whether Barbara has harmed Billy.

In response, I contend that this fact really *doesn't* matter to our commonsense judgments of harm, and that it therefore shouldn't matter to our commonsense judgments of harming. Consider, for example, someone who has a genetic disease like cystic fibrosis. If you were asked whether such a condition were harmful, it is likely that you would answer *yes*, and that your reasoning would involve a comparison between how life is for the individual with the disease and how life would be if she didn't have it. Without cystic fibrosis, the individual could fall asleep at night instead of continually struggling to breathe. She could attend more social events instead of winding up in the hospital. She could eat the foods she wanted to eat without getting sick. And so on. These comparisons inform the judgment that cystic fibrosis is, after all, a *disease*, and they help to justify the resources we devote to finding more effective treatments. The fact that the sufferer would not have existed if the cause of her having the disease (namely, the event of her parents' procreating) had not occurred does not seem to matter to how harmful we think the disease is or how desirable it would be to find a cure.

More generally, many philosophers hold that well-being has two kinds of value. It has *personal value*, or value for the individual whose well-being it is, as well as *impersonal value*, or value that is distinct from the value it has for any individual.<sup>16</sup> The concept of harm is linked to the *personal value* of well-being. That is, harm is always harm *for* someone.<sup>17</sup> That is why it is

<sup>16</sup> A theory is committed to the existence of impersonal value if it holds, for example, that a state of affairs containing five people with headaches is *ceteris parabis* worse than a state of affairs containing five different people, only one of whom has a headache. The former state of affairs is not worse for any particular individual, so if it is worse at all, the worseness must be grounded in impersonal value. Utilitarianism, Rossian pluralism, and any other normative theories that enjoin us consider aggregated well-being appear to be committed to the claim that well-being has impersonal value.

<sup>17</sup> Indeed, the notion that consequentialists are generally concerned with impersonal value, rather than personal value, can explain why many consequentialists are reluctant to acknowledge either the existence or importance of harm at all. See Bradley (2012), who argues that our moral theories should do away with the concept of harm, and Norcross (2005), who writes, “[F]or the purposes of ethical theorizing, harm and benefit do not have the kind of metaphysical grounding required to play fundamental roles in ethical theory, nor do judgments of harm

not *ad hoc* to assess harm from the perspective of the individual who suffers it. In comparing the harm to its absence, we compare worlds where she suffers the harm to worlds where she *exists* but does not suffer it. We cannot assess the personal value of alterations in someone's well-being by making comparisons to worlds where the individual does not exist.

### **3. Revisiting the Problem Cases**

In the previous section I introduced the existence account of harming, defended its tenets, and showed how the account can vindicate the harm-based approach to the non-identity problem. In this section I will first explain how the existence account of harming gets the right results in the cases that created trouble for the non-comparative account of harming. I will also consider another objection to the existence account. In response to this objection, I will defend a principle about the strength of the reason against harming. That principle, when conjoined with the existence account, will help diffuse the objection.

Here is why the existence account of harming is immune to the problems illustrated by Dim Vision, Death, and Loss of Fortune. Recall that in Dim Vision, Dr. Smith causes Jones to be in a state of dim vision. Such a state appears to be bad for Jones, so if the non-comparative account of harming is true, then Dr. Smith harms Jones. However, the existence account of harming does not imply that Dr. Smith harms Jones. Consider the following counterfactual:

- (J) If it were true that both Jones existed *and* he did not have dim vision, then Jones would be better off in some respect.

J states a necessary condition for dim vision's being a harm for Jones. However, J is false. If Jones were to exist without dim vision, then Jones would be completely blind; that would result

---

or benefit make any distinctive contributions to reasons for action" (p. 171 – 172).

in his being *worse off* in some respect. The state that Dr. Smith causes Jones to be in fails to satisfy a necessary condition for being a harm, and since it is not a harm, Dr. Smith does not harm Jones in causing it.

This result may seem worrisome, for as I noted in Section 1, the state of dim vision *appears* to be bad for Jones. However, it is easy to explain away the appearance of badness. What is *actually* bad for Jones is not dim vision, but the state of being unable to see very well. And sure enough, the existence account of harming implies that the state of being unable to see very well *is* a harm for Jones. Consider the following counterfactual:

(W) If it were true that both Jones existed *and* he were *not* *unable* to see very well, then Jones would have a higher level of lifetime well-being.

W is true. To say that Jones is *not* *unable* to see very well is to say that Jones *is* able to see very well. And if Jones were to see very well, then he would indeed be better off in some respect. The existence account of harming thus implies that although Dr. Smith caused dim vision, dim vision is not a harm. On the other hand, being unable to see very well *is* a harm, but Dr. Smith did not cause it.

The existence account of harming is also immune to the objection illustrated by Death. Recall that in Death, Sam dies instantly. The event of his death does not result in a state that is bad for Sam, for the state that results is not a state *in which Sam exists*. That is why the non-comparative account of harming cannot explain why the train accident harms Sam.

However, according to the existence account of harming, the state of affairs that is the standard for assessment is not the actual state of affairs in which Sam no longer exists, but rather the possible state of affairs in which Sam exists without having been killed by the train. The question is whether, in *that* state of affairs, Sam is better off in some respect. And since in that

state of affairs, Sam would live happily and healthfully past the age of 25, he *would* be better off in many respects. Thus, the existence account of harming implies that it is a harm for Sam that he died when he did and that, in causing that harm, the train harmed him.

Finally, the existence account of harming is immune to the objection illustrated by Loss of Fortune. Recall that in Loss of Fortune, Jeeves suffers brain damage that results in his having average intellectual abilities. The proponents of the non-comparative account of harming were at a loss to explain why having average intellectual abilities was a bad state for Jeeves, so they could not use that account to explain why the stroke harmed Jeeves. However, the existence account of harming implies that having average intellectual abilities is a harm for Jeeves, for the following counterfactual is true:

- (I) If it were true that both Jeeves existed *and* he did not have average intellectual abilities, then Jeeves would be better off in some respect.

I is true because if Jeeves did not have average intellectual abilities, then he would not have had the stroke, and so he would have retained his extraordinary intellectual abilities. The existence account of harming implies that the stroke harmed Jeeves.

At this point, one might raise the following objection to my view. As the example of being unable to see very well illustrates, the existence account of harming is compatible with the claim that states of affairs that are either positively described (being blind) or negatively described (*not* being able to see very well) can be harms. But if we allow states of affairs that are negatively described to be harms, then presumably, not having almost *anything* good would qualify as a harm. For example, it is a harm for me that I do not have telekinetic abilities, since if both (1) I existed and (2) I did *not not* have telekinetic abilities, I would be better off in some respect. Yet it is not a harm for me that I lack telekinetic abilities.

I have two responses to this objection. First, I contend that the intuition that not having telekinetic abilities is *not* a harm is not very robust, and consequently, we should disregard it, since we otherwise have strong reason to accept the existence account of harming. The same is true for our intuitions about the lack of other goods. Not having wings, not being able to live to the age of 200, not having the intellectual abilities of a telepathic super genius: these are harms for most, if not all, of us.<sup>18</sup> Admittedly, I noted in Section 1 that it does not *seem* to be bad for humans who are not Jeeves to lack extraordinary intellectual abilities. I also noted that, intuitively, a mother bat does not harm her baby when she causes him to exist in a state such that he is unable to see. I now claim that such intuitions are misleading: we have a principled reason to hold both that it *is* a harm for many people to lack the intellectual abilities of Jeeves and that a mother bat *does* harm her baby when she brings him into a typical bat existence. Nevertheless, we can still claim that any *reason* against what the mother bat does, or against what any human parents would do, were they to bring into existence an individual of ordinary intellectual ability, would be vanishingly weak. The justification for these claims comes from the principle that I alluded to at the start of this section.

Before I state the principle, I will introduce the following definitions:

*Inevitability/Avoidability* (def.): For any state of affairs, A, A is *more inevitable (less avoidable)*, the more the world would be different if A did not obtain, and A is *less inevitable (more avoidable)*, the less the world would be different if A did not obtain.

'Inevitability' and 'avoidability' are terms of art. In ordinary language, we tend not to think that 'inevitable' and 'avoidable' admit of degree. However, as I am using these terms, they do.

<sup>18</sup> Cf. Thomas Nagel, who writes, "The question is whether we can regard as a misfortune any limitation, like mortality, that is normal to the species. .... A man's sense of his own experience ... does not embody this idea of a natural limit. His existence defines for him an essentially open-ended possible future, containing the usual mixture of goods and evils that he has found so tolerable in the past. ... Viewed in this way, death, no matter how inevitable, is an abrupt cancellation of indefinitely extensive possible goods. Normality seems to have nothing to do with it, for the fact that we will all inevitably die in a few score years cannot by itself imply that it would not be good to live longer" (1970, p. 80).

To see how these terms can be used, consider the actual state of affairs in which Barack Obama wins the 2008 Presidential election. This state of affairs is slightly more inevitable than the state of affairs in which Mitt Romney wins the 2008 election, for the world in which Romney wins is slightly different from the actual world. The state of affairs in which Obama wins is also *much* more inevitable than the state of affairs in which a goat wins the 2008 election. The world in which a goat wins is very different from the actual world. Conversely, the world in which a goat wins is highly avoidable, the world in which Romney wins is less avoidable, and the world in which Obama wins is the least avoidable of the three states of affairs.

With these concepts of inevitability and avoidability in hand, consider the following principle:

*The Inevitable Harming Principle:* Other things being equal, the reason against harming an individual is weaker, the more inevitable the harm.<sup>19</sup>

We can now apply this principle to the cases of Jeeves and the bat. For a bat, being unable to see is a harm that is highly inevitable. That is because, in a world where a bat could see, there would be extraordinary developments in genetic engineering, and such a world is very different from our own. Given how inevitable blindness is for a bat, the inevitable harming principle implies that, other things being equal, the reason against causing a bat to be born blind is much weaker than the reason against causing a human to be born blind.

Moreover, not having extraordinary intellectual abilities is more inevitable for the average person than it is for Jones. Therefore, according to inevitable harming principle, the reason against causing the average person not to have extraordinary intellectual abilities is much weaker than the reason against causing Jones not to have extraordinary intellectual abilities.

---

<sup>19</sup> This is not intended to be a complete account of the strength of the reason against harming. In other work, I defend a longer list of principles that determine the strength of such a reason; see my unpublished manuscript.

Compare two cases that are variations on Case 1 and Case 2:

*Case 3.* Alice is already pregnant with the fetus who will become Jeeves. She decides to take a brain damage-causing drug that will cause Jeeves to have average intellectual abilities instead of extraordinary ones. After he is born, Jeeves develops average intellectual abilities.

*Case 4.* Barbara is using *in vitro* fertilization. She would like to have a child who has average intellectual abilities, so she selects an embryo on this basis. After he is born, her child, whom she names Reeves, develops average intellectual abilities.

Having average intellectual abilities is more inevitable for Reeves than it is for Jeeves. Given that other factors are equal between the cases, the inevitable harming principle implies that the reason against what Alice does is stronger than the reason against what Barbara does.<sup>20</sup> Since we can distinguish between the strength of reasons in this way, we need not bother with further distinctions implying that blindness is a harm for a human but not for a bat, or that not having extraordinary intellectual abilities is a harm for Jeeves but not for Reeves.<sup>21</sup>

Notice that the inevitable harming principle adds a nuance to what the existence account of harming says about the non-identity problem. In Section 1, I compared Alice's action in Case 1 to Barbara's action in Case 2. I said that part of the problem was to justify the intuition that Barbara's action is objectionable in much the same way that Alice's action is objectionable. Now, conjoined with the inevitable harming principle, the existence account of misfortune does, indeed, imply that Barbara's action is objectionable in much the same *way* that Alice's is: both actions are instances of harming. Nevertheless, the inevitable harming principle also implies

---

<sup>20</sup> Indeed, for any action that causes an independently existing person to suffer a misfortune, it is likely that, according to inevitable harming principle, the reason against the action is strong. Consider a typical case in which an action causes a person who would exist regardless of the action to suffer a harm. In such a case, the harm is highly avoidable: if the harm did not obtain, the only thing that would be different would be that the action was not performed. Since the harm is highly avoidable, the reason against the action (other things being equal) is strong.

<sup>21</sup> We might also think that blindness is less of a harm for a bat than it is for a human because being able to see would not improve a bat's well-being as much as it would improve a human's well-being. The existence account of harming is compatible with this explanation.

Barbara's action is objectionable to a lesser *degree* than Alice's is. After all, Billy's poor health is much more inevitable than Alex's, and so there is less of a reason for Barbara to refrain from causing it. I take it that this is the right result. Most people I present the cases to have the intuition that what Barbara does wrongs Billy, but that it doesn't wrong Billy to the same degree that Alice's action wrongs Alex.<sup>22</sup>

Finally, notice that the inevitable harming principle does not have objectionably arbitrary implications about what the biologists should do in Goat. Recall that in that case, the biologists were wondering whether it would be morally permissible to let the embryos in the middle of the continuum between humans and goats develop into adults. Since the existence account of harming does not endorse a substantive account of well-being, it does not answer the question about whether life as a half-human, half-goat is bad for the individual living it. Nevertheless, if such a life *is* bad for such an individual, the existence account of harming implies that the biologists will be harming the resultant individuals if they cause the embryos to develop. Moreover, the inevitable harming principle implies that the reason against causing such harm will be somewhat consistent in strength across most of the continuum, since existence as a part-human, part-goat will be equally inevitable for each individual on the continuum, excluding those on the very ends.

I conclude that the existence account of harming can not only vindicate the harm-based approach to the non-identity problem, but it can also serve as a plausible account of harming in its own right.

## Bibliography

Adams, Robert, "Must God Create the Best?" *The Philosophical Review*, vol. 81, no. 3 (1972), pp. 317–332.

---

<sup>22</sup> Justin Weinberg (2013) also defends this intuition.

- Boonin, David, "How to Solve the Non-Identity Problem," *Public Affairs Quarterly*, vol. 22, no. 2 (2008), pp. 127 – 157.
- . (2015) *The Non-Identity Problem and the Ethics of Future People* (Oxford: Oxford University Press).
- Bradley, B. (2012) "Doing Away with Harm" *Philosophy and Phenomenological Research* 85(2): 390 – 412.
- Feinberg, Joel, *Harm to Others* (Oxford: Oxford University Press, 1984).
- Gardner, M. "On the Strength of the Reason Against Harming." (unpublished manuscript)
- Greene, Mark and Steven Augello, "Everworse: What's Wrong with Selecting for Disability?" *Public Affairs Quarterly*, vol. 25, no. 2 (2011), pp. 131 – 139.
- Hanser, Matthew, "The Metaphysics of Harm," *Philosophy and Phenomenological Research*, vol. 77, no. 2 (2008), pp. 421 – 450.
- , "Harming and Procreating," in *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*, ed. Melinda Roberts and David Wasserman (Dordrecht: Springer, 2009), pp. 179 – 199.
- Harman, Elizabeth, "Can We Harm and Benefit in Creating?" *Philosophical Perspectives* 18 (2004), pp. 89 – 113.
- , "Harming as Causing Harm," in *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*, ed. Melinda Roberts and David Wasserman (Dordrecht: Springer, 2009), pp. 137 – 154.
- Kavka, Gregory, "The Paradox of Future Individuals," *Philosophy & Public Affairs*, vol. 11 (1982), pp. 93–112.
- Klocksiem, J. "A Defense of the Counterfactual Comparative Account of Harm" *American Philosophical Quarterly* 49(4), pp. 285 – 300.
- McMahan, Jeff, "Cognitive poor health, Misfortune, and Justice," *Philosophy & Public Affairs*, vol. 25, no. 1 (1996), pp. 3 – 35.
- Nagel, T. (1970) "Death," *Noûs* 4(1): 73 – 80.
- Parfit, D. (1976) "On Doing the Best for Our Children," in *Ethics and Population*, ed. Michael Bayles (Cambridge, Mass.: Schenkman), pp. 100 – 102.
- , (1987) *Reasons and Persons* (Oxford: Clarendon Press).
- Pitcher, George, "The Misfortunes of the Dead," *American Philosophical Quarterly*, vol. 21, no. 2 (1984), pp. 183 – 188.
- Roberts, M. A. (2013) "The Nonidentity Problem," *The Stanford Encyclopedia of Philosophy* (Fall 2013 Edition), Edward N. Zalta (ed.), URL: <<http://plato.stanford.edu/archives/fall2013/entries/nonidentity-problem/>>.
- Schwartz, Thomas, "Obligations to Posterity," in *Obligations to Future Generations*, ed. Richard Sikora and Brian Barry (Philadelphia: Temple University Press, 1978), pp. 3 – 13.
- Shiffrin, Seana, "Wrongful Life, Procreative Responsibility, and the Significance of Harm," *Legal Theory*, vol. 5 (1999), pp. 117–48.
- Thomson, Judith "More on the Metaphysics of Harm," *Philosophy and Phenomenological Research*, vol. 82, no. 2 (2010), pp. 436 – 458.
- Weinberg, J. (2013) "Non-Identity Matters, Sometimes" *Utilitas* 26(1): 1 – 11.
- Woollard, Fiona, "Have We Solved the Non-Identity Problem?" *Ethical Theory and Moral Practice*, vol. 15, no. 5 (2012), pp. 677 – 690.