

David Boonin on the Non-Identity Argument: Rejecting the Second Premise¹

MOLLY GARDNER

Bowling Green State University

ABSTRACT

According to various “harm-based” approaches to the non-identity problem, an action that brings a particular child into existence can also harm that child, even if his or her life is worth living. In the third chapter of *The Non-Identity Problem and the Ethics of Future People*, David Boonin surveys a variety of harm-based approaches and argues that none of them are successful. In this paper I argue that his objections to these various approaches do not impugn a harm-based approach that Boonin does not consider, an approach I call the “existence solution to the non-identity problem”. I also argue that the existence solution is more plausible than Boonin’s own proposed solution.

Keywords: non-identity problem, David Boonin, harm-based approach, existence solution, causal selection, mere condition, backtracking counterfactuals.

1. INTRODUCTION

One strategy for solving the non-identity problem is to hold that in some cases, an action that brings a particular child into existence can also harm him or her, even if the child’s life is worth living. This approach, which I will call the “harm-based” approach, is the focus of the third chapter of David Boonin’s (2014) book, *The Non-Identity Problem and the Ethics of Future People*. In the chapter, Boonin considers versions of the harm-based approach developed by Matthew Hanser (2008, 2009), Elizabeth Harman (2004, 2009), Seana Shiffrin (1999), and others, and he argues that either these views fail to deliver a sufficiently robust solution to the non-identity problem or else they lack plausibility (which Boonin refers to as “modesty”).

¹ For helpful feedback on this paper, I thank Eden Lin and two anonymous referees for this journal.

He concludes that *no* version of the harm-based approach will succeed, for the weaknesses of the views he criticizes will afflict all other versions of the approach.

Has Boonin really identified fatal flaws in the views he considers, and do these flaws generalize to all other versions of the harm-based approach? I agree that some of Boonin's objections are persuasive: Boonin has, indeed, shown that some versions of the harm-based approach are either immodest or insufficiently robust. However, I disagree with Boonin's claim that these objections generalize to *all* versions of the harm-based approach. In what follows, I will argue that a particular version of the harm-based approach that Boonin does not consider—one that features the *existence account of harming*²—does, in fact, deliver a robust solution to the non-identity problem. Moreover, the costs of what I will call “the existence solution” are minimal: it requires less bullet-biting than other versions of the harm-based approach, and indeed, less bullet-biting than Boonin's own view.

To make my case, I will first provide a more precise characterization of both the non-identity problem and the harm-based approach to solving it. I will then explain what Boonin means when he objects to other versions of the harm-based approach on the grounds of either immodesty or a lack of robustness. In the fourth section, I will explain both the existence account of harming and the existence solution that it supports, and I will argue that the existence solution is neither insufficiently robust nor immodest. Finally, I will argue that the implications of Boonin's own solution to the non-identity problem are less plausible, on balance, than are those of the existence solution.

2. THE NON-IDENTITY PROBLEM AND THE HARM-BASED APPROACH

The non-identity problem arises in cases where the existence of an individual whose life is worth living depends counterfactually upon an action that appears to be objectionable in virtue of some negative consequence the action has for the individual's well-being. To illustrate the problem, Boonin constructs a case in which a doctor tells a woman named Wilma that if she conceives a child now, the child will be blind, but if she takes a pill once a day for two months and conceives after that, her child's vision will be normal. Importantly, the blind child who would be born now

² I develop the existence account of harming more fully in Gardner (2015) and Gardner (2017).

(“Pebbles”) is not identical to the child who would be born two months from now (“Rocks”)—hence the term “non-identity”.³ The doctor tells Wilma that the pills are easy to take and have no side effects. Nevertheless, Wilma decides against taking the pills, and she conceives Pebbles instead of Rocks. Pebbles’ blindness is a significant detriment to her quality of life, but on the whole, her life is worth living.

To see how this case generates a philosophical problem, consider the tension between various intuitions people might have about the case. On the one hand, Pebbles is worse off than Rocks would have been had Wilma acted differently, and this consideration might suggest that there was something wrong with Wilma’s action. But on the other hand, Pebbles is no worse off than Pebbles would have been had Wilma acted differently; after all, if Wilma had not conceived now, *Pebbles* would not have been at all. This second consideration lends support to the notion that Wilma’s action was unobjectionable. Boonin attempts to capture the tension between these considerations with what he calls the “non-identity argument”, an argument whose five premises are highly plausible in themselves, but whose conclusion is so counterintuitive that Boonin refers to it as “the Implausible Conclusion”:

- (P1) Wilma’s act of conceiving now rather than taking a pill once a day for two months before conceiving does not make Pebbles worse off than she would otherwise have been.
- (P2) If A’s act harms B, then A’s act makes B worse off than B would otherwise have been.
- (P3) Wilma’s act of conceiving now rather than taking a pill once a day for two months before conceiving does not harm anyone other than Pebbles.
- (P4) If an act does not harm anyone, then the act does not wrong anyone.
- (P5) If an act does not wrong anyone, then the act is not morally wrong.

Therefore,

- (C) Wilma’s act of conceiving Pebbles is not morally wrong.

As Boonin notes, to solve the non-identity problem, we must either reject at least one of the plausible premises or accept the Implausible Conclusion. In the various chapters of his book, Boonin considers attempts to reject each of the five premises. He concludes that none of these attempts succeed, and he ends the book by endorsing the Implausible Conclusion.

³ The term “non-identity problem” was coined by Derek Parfit (1984) to describe a problem that appears to have been independently discovered by Parfit (1976), Thomas Schwartz (1978), and Robert Adams (1979).

For my purposes, the relevant chapter is the third one, in which Boonin criticizes various attempts to reject P2. What I'm calling the "harm-based" approach to the non-identity problem can be understood as any approach that tries to solve the non-identity problem by rejecting P2. To reject P2 is to maintain that an action can harm someone even if it doesn't make the victim worse off than he or she would otherwise have been. Those who reject P2 usually appeal to some set of harm-based principles or some account of harming in order to justify their view. In Section 3, I will defend the account of harming that I believe can *best* justify the rejection of P2, namely, the existence account. First, however, I will review the two main objections that Boonin raises against other versions of the harm-based approach.

3. ROBUSTNESS AND MODESTY

In organizing his objections to other versions of the harm-based approach, Boonin appeals to what he calls "the robustness requirement" and "the modesty requirement": two requirements that a successful solution to the non-identity problem must meet. He claims that all versions of the harm-based approach violate either the robustness requirement, the modesty requirement, or both; this is why he believes that no version of the harm-based approach will successfully solve the non-identity problem. In this section, I will explain Boonin's account of each requirement in turn.

According to Boonin, a principle that purports to solve the non-identity problem by rejecting a premise of the non-identity argument meets the *robustness* requirement only if it is "strong enough to warrant rejecting any weakened version of the premise that would still be strong enough to generate the unqualified version of the Implausible Conclusion" (22). There are two things to note about this account of robustness. First, the phrase, "unqualified version of the Implausible Conclusion" is supposed to refer to the conclusion that Wilma's act of conceiving Pebbles—in the case as originally described—is not morally wrong. A *qualified* version of that conclusion would involve adding new facts to the case, such as the fact that by conceiving Pebbles, Wilma violates a just law or breaks a promise. If we added the latter fact to the case, the qualified conclusion would be that Wilma's act of conceiving Pebbles *and thereby breaking a promise* is not morally wrong. It is easy to undermine such a qualified conclusion, but as Boonin points out, doing so does not really help us solve the non-identity problem.

The second thing to note about Boonin's account of robustness is his reference to a "weakened version of the premise". Recall that premise two

of the non-identity argument says that if A's act harms B, then A's act makes B worse off than B would otherwise have been. Here is a weakened version of that premise:

(W2) If Wilma's act (in the case as originally described) harms Pebbles, then Wilma's act makes Pebbles worse off than Pebbles would otherwise have been.

Boonin's point is that it might be possible to find counterexamples to P2 that would not be counterexamples to W2. But since W2 is still strong enough, on its own, to generate the unqualified Implausible Conclusion, a counterexample to P2 that did not undermine W2 would not really solve the non-identity problem. Thus, a harm-based solution to the non-identity problem satisfies the robustness requirement if and only if it undermines W2.

In addition to the robustness requirement, Boonin's objections to the harm-based views also appeal to what he calls "the modesty requirement". According to Boonin, a proposed solution to the non-identity problem will fail the modesty requirement if its implications are less plausible than the Implausible Conclusion. He writes, "The modesty requirement constitutes a reasonable constraint on a satisfactory solution to the non-identity problem because considerations about intuitive plausibility are precisely what give rise to the problem in the first place" (23). Thus, to meet the modesty requirement, a solution based on the existence account of harming (or any other view) must have implications that are more plausible than the claim that Wilma's act of conceiving Pebbles is not morally wrong.

As I noted above, Boonin's main objections to various versions of the harm-based approach are that each version violates either the robustness requirement, the modesty requirement, or both. Thus, in order to show that the existence solution to the non-identity problem is impervious to Boonin's main objections, I will need to show that it satisfies both requirements. In the next section, I will explain the existence solution to the non-identity problem and I will argue that it does, indeed, satisfy both.

4. SOLVING THE NON-IDENTITY PROBLEM WITH THE EXISTENCE ACCOUNT OF HARMING

I will introduce the existence account of harming by contrasting it with the second premise of the non-identity argument. Recall that according to P2, if A's act harms B, then A's act makes B worse off than B would otherwise have been. When we interpret this principle, it is natural to construe the word 'makes' as a causal verb: another way of stating P2 is to say that a

harmful action always *causes* the victim to be worse off than he or she otherwise would have been. And indeed, Boonin often uses causal language and counterfactual language interchangeably. Thus, he seems committed to the view that the following sentence implies that A causes B:

(D) If A hadn't happened, then B wouldn't have happened.

However, the existence solution expressly rejects the notion that the D states a sufficient condition for causation. According to the existence solution, causation may be reducible to counterfactual dependence, but not in a way that makes the relation expressed by D sufficient for causation: the relationship between causation and counterfactual dependence is more complicated, and the two concepts play distinct roles in the solution to the non-identity problem.⁴ Here is the full 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, if and only if

- (i) There is an essential component of T that is a condition with respect to which S can be intrinsically better or worse off; and
- (ii) If S existed and T had not obtained, then S would be better off with respect to that condition.⁵

This account has at least three important features. First, it is compatible with the rejection of P2 in the non-identity argument. The existence account of harming does not imply that an action harms a victim only if the action makes the victim worse off than he or she would have been, had the *action* not been performed. Instead, it implies that an action harms a victim only if it makes the victim worse off than he or she would have been, had the victim existed without the upshot of the action. The existence account of harming thereby locates harm in the *effects* of various actions, and not in the *causes* of those effects.⁶

Second, to assess whether an effect qualifies as a harm, the account does not ask us to compare a world in which the victim suffers the alleged harm to a possible world in which the victim does not exist at all. Instead, the account asks us to compare the world containing both the victim and

⁴ I say more about the relationship between counterfactual dependence and causation in Gardner (forthcoming).

⁵ This formulation of the account comes from Gardner (2015).

⁶ If the reader is still uncertain what this means, don't despair. Very shortly, I will illustrate how the view works by applying it to the case of Wilma and Pebbles. For more on the advantages of an "effect-relative" account of harming over an "action-relative" account, see Gardner (2017).

the alleged harm to a world where the victim exists without the alleged harm; we then check to see whether, in the latter world, the victim is better off. Such a comparison is legitimate because it is metaphysically possible for someone to exist without, say, being blind, even if, as a matter of fact, she *would* not have existed if the cause of her blindness hadn't obtained. Even if your formative years are essential to your identity, such that you could not have existed if you had not been born blind, it is still metaphysically possible to *become* sighted at the age of, say, 15. So there is a possible world where you exist and are sighted, and we can compare your well-being in that possible world to your well-being in the world in which you are blind. Such a comparison is appropriate for ascertaining whether a state of affairs qualifies as a harm because a harm is the sort of thing you would be better off without. If we don't consider the possible worlds where you exist in the absence of the alleged harm, it is impossible to determine whether you would be better off without it.⁷

A third important feature of the existence account of harming is that it yields a solution to the non-identity problem that is impervious to the two main objections that Boonin raises against other versions of the harm-based approach. To show that this is so, I will explain how the existence solution satisfies both the robustness requirement and the modesty requirement, in turn.

Recall that a solution to the non-identity problem satisfies the robustness requirement if and only if it undermines W2, the claim that if Wilma's act (in the case as originally described) harms Pebbles, then Wilma's act makes Pebbles worse off than Pebbles would otherwise have been. Thus, I need to show that the existence account of harming implies, not only that an act can harm someone without making the victim worse off, but that *Wilma's* act, in the case as originally described, harms Pebbles without making her worse off. I will show this by arguing that (a) there is a state of affairs that instantiates a comparatively poor showing along some intrinsic dimension of Pebbles' well-being and (b) Wilma's act causes that state of affairs.

To see why there is a state of affairs that instantiates a comparatively poor showing along some intrinsic dimension of Pebbles' well-being, recall the original description of the case involving Wilma and Pebbles. I noted there that Pebbles' blindness is a significant detriment to her quality of life. This stipulation about the case (which is also featured in Boonin's description of the case) means that for Pebbles, things are going comparatively worse along some dimension of her well-being. (If her well-being has only one dimension, then things are going comparatively worse

⁷ For more on why we should build the assumption that S exists into the antecedent of the relevant counterfactual, see Gardner (2015: 436 – 437).

for her, full stop.) What, exactly, is measured by that dimension of well-being depends on which theory of well-being is correct; it may be that Pebbles is doing poorly with respect to her capabilities, her happiness, the satisfaction of her desires, or some other objective list of goods. Indeed, to more precisely describe the state of affairs that constitutes the harm for Pebbles, we would need to know which theory of well-being is correct. For example, if an objective list theory were true and if vision were on the objective list of goods, then the harm for Pebbles would be *the state of affairs in which she is blind*. But we do not need to describe the state of affairs so precisely in order to know that there is such a state of affairs. All we need to know is that there is some well-being-related state of affairs such that, if Pebbles existed and that state of affairs had not obtained, Pebbles would be faring better. And since we do know this, conditions (i) and (ii) in the above definition of harm are both satisfied.

We have so far established that Pebbles suffers a harm. The next question is whether Wilma's act causes that harm. Does it? Someone might worry that we need to describe the state of affairs that is the harm more precisely in order to know whether Wilma's act causes that harm. However, this is not so: in the case as described, the harmful state of affairs is closely connected to (if not *identical* to) the state of affairs in which Pebbles is blind. Therefore, if Wilma's act causes the state of affairs in which Pebbles is blind, then *by causing* that state of affairs, her act also causes the state of affairs that is a harm for Pebbles. Thus, in order to show that Wilma's act harms Pebbles, we need only to show that Wilma's act causes the state of affairs in which Pebbles is blind.

Notice that a paradigmatic way of causing a state of affairs in which someone is blind is to blind someone after they have already been born. For example, suppose that Ted throws acid in Fred's eyes, and from that point on Fred is blind. This is a clear case in which (a) Fred suffers a harm and (b) Ted caused that harm.

Wilma's act is different from Ted's. Instead of inducing a certain biological property in an individual who already exists, Wilma's act *brings into existence* an individual with that property. However, this difference is not as significant as it might initially appear to be. That is because the existence account of harming does not say that harming is a matter of causing someone to have a certain property. Instead, it says that harming is a matter of causing a particular state of affairs. And although the case of Ted and Fred shows that causing someone *to have* a certain property is *one* way of causing a particular state of affairs, it is not the only way. There are, in other words, at least two ways to cause the harm of blindness: you can cause someone who already exists to be blind, or you can cause someone

to exist who *will* be blind. Wilma causes the existence of Pebbles, who will be blind, and so she causes a state of affairs that is a harm for Pebbles. This means that the definition of harming provided by the existence account of harming is also satisfied: Wilma's act harms Pebbles. Because the existence account of harming yields this result, Boonin's robustness requirement is satisfied.

What about the modesty requirement? Recall that to meet the modesty requirement, a solution to the non-identity problem must not have implications that are less plausible than the Implausible Conclusion. Boonin does not consider the existence solution in his chapter, so I cannot be entirely sure which implications of the existence solution he would select for entry into a showdown with the Implausible Conclusion. Nevertheless, the two implications of the existence solution that Boonin and other opponents would probably find the most objectionable are the following:

- (1) Counterfactual dependence is not a sufficient condition for causation; and
- (2) Some so-called "backtracking counterfactuals" are true.

In the remainder of this section, I will explain why the existence solution has each of these implications and then argue that such implications are not as bad as they might initially seem.

First, however, I want to make an observation. In his critique of other harm-based views, the implications that Boonin objects to are not pure metaphysical principles in the way that (1) and (2) are. Instead, they are implications about concrete cases. Some of the implications that Boonin objects to, for example, are the following:

- (3) When I prevent you from receiving the medicine you need, I don't harm you (60).
- (4) When I "shoot you in self-defense or as part of the prosecution of a just war", I don't harm you (63).
- (5) When a doctor performs a procedure that causes a patient named Ray to regain his vision for forty years and then lose it again, the doctor wrongfully harms Ray (68).
- (6) Every act of conception harms the person who is conceived (70).

The relative concreteness of these claims makes it easier for Boonin to argue that they are less plausible than the Implausible Conclusion. This is because Boonin often appeals to what his colleagues and students actually think in order to establish the comparative plausibility of various

propositions (25). And in my experience, students, in particular, are relatively confident when they form judgments about concrete cases of stealing medicine, shooting people, operating on people, conceiving children, and so on. However, I find that students tend to be less confident when they are asked how plausible claims like (1) and (2) are; there is less of a basis in what we might call “commonsense metaphysics” for either affirming or rejecting (1) and (2) directly. So to test (1) and (2) against common sense, we need to see how these abstract claims function within broader metaphysical theories that, through complicated chains of entailment, might eventually support or undercut our more easily testable, commonsense intuitions. I believe that (1) and (2) fit extraordinarily well within a broader metaphysical picture that supports a great deal of our commonsense intuitions about causation, harm, and wrongdoing. I cannot fully support such a claim in this article, but I will present some of the main reasons I believe this is so. At this point, however, I wish to simply to acknowledge the fact that the ostensibly implausible implications of the existence solution are slightly more abstract than the ostensibly implausible implications of other harm-based views.

Here is why the existence solution is committed to (1), the claim that counterfactual dependence is not a sufficient condition for causation. According to the existence solution, Wilma harms Pebbles by conceiving her, and this helps explain why Wilma’s action was morally wrong. However, if counterfactual dependence *were* a sufficient condition for causation, then *every* parent would harm his or her child by bringing that child into existence. After all, we all suffer numerous harms throughout our lives: we get sick, we get injured, we suffer distress, and of course, we die. All of the corresponding harms depend counterfactually on the proposition that our parents conceived us. If our parents had *not* conceived us, we would not have suffered such harms. Thus, if counterfactual dependence were sufficient for causation, there would be little point in focusing on the harm associated with Pebbles’s blindness—just one of the *myriad* harms that Wilma causes Pebbles, and probably not nearly the worst—and supposing that our knowledge of this one harm could, by itself, settle the question of whether Wilma acted wrongly.

Instead of endorsing counterfactual dependence as a sufficient condition for causation, the existence solution affirms a distinction between a cause and a mere condition. According to the existence solution, there is a good reason for Wilma to focus on Pebbles’ blindness, rather than all the other harms that Pebbles will suffer: namely, Wilma causes the state of affairs in which Pebbles is blind, but her action is merely the condition of those other harms. From Wilma’s perspective, Pebbles’ blindness generates

a relatively strong reason against conceiving Pebbles, and those other harms do not. And if, as seems to be true in the case Boonin constructs, this reason against conceiving Pebbles is not outweighed by countervailing reasons, then conceiving Pebbles is wrong.

That is why the existence solution is committed to the claim that counterfactual dependence is not sufficient for causation. But is this commitment more implausible than the Implausible Conclusion? *It would* be more implausible if there were no defensible principles of causation that could accommodate such a commitment. But fortunately for the existence solution, there are coherent principles of causation that reject counterfactual dependence as sufficient for causation and instead support a distinction between a cause and a mere condition.⁸ Moreover, there is good reason to accept such causal principles; they cohere remarkably well with the judgements about fault and blame that we make in law and everyday morality.⁹ We don't, for example, blame our parents for our deaths (unless they do, in fact, kill us); we don't blame the fall of Rome on the meteor that wiped out the dinosaurs; and we don't attribute the election of Donald Trump to the existence of Earth's magnetic field. Thus, insofar as common sense morality has a role to play in vindicating Boonin's view, it can play that role equally well or better in vindicating the distinction between a cause and a mere condition.

As I noted above, the second implication that I believe opponents of the existence solution would object to is the claim that some so-called "backtracking counterfactuals" are true. To understand what a backtracking counterfactual is, consider two events, A and B, and suppose that B happens later in time than A. A foretracking counterfactual—the kind I have rejected as a sufficient condition for causation—takes this form:

(O) If A hadn't happened, then B wouldn't have happened.

By contrast, a backtracking counterfactual takes this form:

(B) If B hadn't happened, then A wouldn't have happened.

Thus, backtracking counterfactuals express a way of reasoning backwards: we reason that if a later event had not happened, an earlier event would not

⁸ For some causal principles that support a distinction between a cause and a mere condition, see Hart and Honoré (1985), Broadbent (2008), Gardner (forthcoming), and McGrath (2005). The existence solution is committed to the claim that at least *one* such account is correct.

⁹ Indeed, Hitchcock (2007) contrasts what he calls a "folk attributive" concept of causation—a concept that includes a distinction between a cause and a mere condition—with both a "metaphysical" concept of causation and a "scientific" concept of causation. He argues that the folk attributive concept and the scientific concept have promise, but he expresses skepticism that we need a theory of the "metaphysical" concept.

have happened. When uttered, many backtracking counterfactuals sound somewhat more awkward than foretracking counterfactuals, and this may be one of the reasons that philosophers tend not to appeal to them. As Alex Broadbent argues, it sounds odd to say, “If I had been richer, I would have chosen a different career”. Nevertheless, he notes that the inference, *itself*, is not odd, for we can make sense of the claim, “For me to have been richer, I would have to have chosen a different career”. Moreover, he notes that not all backtracking counterfactuals sound so odd, even when we leave out constructions like ‘would have to have’; there is nothing objectionable about saying “If the librarian had not gone into the library, she would not have left her fingerprints on the door”¹⁰.

To see why the existence solution is committed to the claim that some backtracking counterfactuals are true, consider the following case:

Dim Vision. Because of retinal damage he suffered long ago, Jones has been blind for many years. Dr. Smith performs an operation that repairs some but not all of the damage. After the operation, Jones can see, but not very well: he has a condition we can call *dim vision*.¹¹

A common intuition about this case is that Dr. Smith does not harm Jones; instead, he benefits Jones by improving his vision. The existence account of harming can justify this intuition if we appeal to the following backtracking counterfactual:

(V) If Jones existed and the state of affairs in which he has dim vision did not obtain, then Dr. Smith wouldn’t have operated and Jones would have been worse off in some respect.

If V is true, then dim vision isn’t a harm for Jones, and Dr. Smith didn’t harm Jones by causing it. But if no backtracking counterfactuals are true, then V *isn’t* true, and the existence account of harming cannot justify the claim that Dr. Smith did not harm Jones. This would be a problem for the existence account of harming, given that there are many cases of helping that bear a structural similarity to Dim Vision, and the existence account would have to hold that all of these were cases of harming, rather than

10 See his unpublished book manuscript, *No Smoke Without Fire*: 18 – 19.

11 This case comes from Hanser (2009) and is also discussed in Harman (2009), Thomson (2010), and Gardner (2015).

helping.¹²

Fortunately, I do not think it is implausible to hold that some backtracking counterfactuals are true. As I noted above, such counterfactuals appear to support many of our inferences, especially those that are inferences to the best explanation. Moreover, there is support for the truth of some backtracking counterfactuals in the metaphysics literature.¹³ To my knowledge, we currently lack a fully developed semantics for evaluating the truth or falsity of backtracking counterfactuals. Nevertheless, we lack an uncontroversial semantics for evaluating the truth or falsity of foretracking counterfactuals as well. This hasn't stopped us from incorporating foretracking counterfactuals into our theories and making intuitive judgments about whether they are true or false, so it should not stop us from appealing to and using our intuitions to evaluate backtracking counterfactuals.

5. A BULLET-BITING SHOWDOWN

Let us take stock. I have so far explained the non-identity problem, the harm-based approach to solving it, and how the existence solution fits into the harm-based approach. I have also reviewed Boonin's two requirements for successfully solving the non-identity problem—the robustness requirement and the modesty requirement—and I have argued that the existence solution to the non-identity problem satisfies both of them. In this final section, I will present some reasons for thinking that the existence solution is a *better* solution than Boonin's own proposal.

As I noted in the introduction, Boonin's own proposal for solving the non-identity problem is to accept the Implausible Conclusion. His reason for doing so is that he believes that every other proposal for solving the

¹² Notice that the backtracking counterfactual we appeal to in Dim Vision has us “imagine away” Dr. Smith's action and then consider how much better or worse off Jones would be. In effect, then, the possible world we consider when we imagine the absence of Jones's new condition is the same possible world we imagine when we employ the principle that serves as P2 in the non-identity argument. This is why, in cases that don't involve either preemption or bringing a person into existence, the existence account of harming delivers all the same verdicts that are delivered by the account of harming that Boonin favors (2014: 53). Importantly, however, the existence account of harming delivers different verdicts in preemption and non-identity cases. To see why the verdict is different in non-identity cases, compare the following two counterfactuals: (1) If Jones existed and didn't have dim vision, Dr. Smith wouldn't have operated; and (2) If Pebbles existed and weren't blind, then Wilma wouldn't have conceived her. Whereas the first counterfactual is true, the second is obviously false: the possibility that Pebbles existed and Wilma didn't conceive her is remote, to say the least.

¹³ See, for example, Elga (2001) and Vihvelin (1994).

problem either doesn't really solve the problem, as advertised, or has implications that are even *less* plausible than the Implausible Conclusion. It is tempting to think, then, that Boonin's endorsement of the Implausible Conclusion is the *least* plausible feature of his view, and that his view as a whole is the *most plausible* of all the approaches to the non-identity problem.

However, as he proceeds towards accepting the Implausible Conclusion, Boonin bites some other bullets along the way. If we count the Implausible Conclusion as one implausible tenet of his view, then his view as a whole includes at least *three* implausible tenets. In all three cases, Boonin provides independent arguments for thinking that the three tenets are not as implausible as they initially seem. However, it's not clear that these other arguments are fully successful. In this final section, I want to explain the two implausible commitments Boonin's view has in addition to the Implausible Conclusion; challenge some of his reasons for holding that these commitments are plausible, after all; and show why the existence solution avoids both of them.

The first bullet Boonin bites has to do with pre-emption cases like the following:

Preemption. Shooter shoots Victim in the abdomen, and although Victim does not die, he must spend the next two weeks in the hospital. If Shooter hadn't shot Victim, then Booter would have, and Victim would have ended up in the hospital with exactly the same injuries.¹⁴

Recall that according to P2, if A's act harms B, then A's act makes B worse off than B would otherwise have been. Because Boonin accepts P2, he is committed to the implausible claim that in *Preemption*, Shooter does not harm Victim.

Boonin accepts this implication of his view and argues that it is more plausible than it might initially appear to be. To show that such a claim is plausible, he constructs a number of thought experiments that he takes to be structurally similar to *Preemption*. He argues that in each case, the relevant action does not harm the victim, and that *if* it does not harm the victim, then Shooter's action in *Preemption* doesn't harm Victim, either. Here is a representative case:

Trolley. Philip is trapped on a runaway trolley. The trolley is about to smash into a brick wall, which will kill him instantly. You can prevent this from happening by pulling a switch before the trolley approaches

14 This case is modeled on the case Boonin calls "The Two Hit Men" (2014: 57).

a branch in the line. If you pull the switch, the trolley will be diverted and will instead smash into a second brick wall, which will also kill Philip instantly and at the exact same moment that the first wall would have killed him. You pull the switch. As a result, Philip is killed by smashing into the second brick wall rather than the first (58).

The problem with Boonin's appeal to Trolley and similar cases is that our intuitions about these cases may not be finely tuned enough to support the judgment that

(A) The action did not harm the victim

rather than

(B) The reason against the action was vanishingly weak.

In other words, both A and B might be reasonable ways to articulate our intuitions about Boonin's cases. However, only A can support Boonin's claim that it is plausible to think that Shooter's action does not harm Victim in Preemption; B cannot. And it is not clear why we should think that A captures our intuition better than B.

Moreover, there are reasons to think that the judgment we should really form about Boonin's cases is B, rather than A. If we conclude that in Preemption, Shooter does not harm Victim, various features of the case are difficult to explain. Why does Victim end up in the hospital for two weeks, if he wasn't harmed? Why is there blood leaking out of his abdomen? Importantly, the claim that Shooter harms Victim is compatible with the claim that the reason against shooting Victim was vanishingly weak: there is always a reason against harming, but the strength of that reason clearly varies from one case to another.¹⁵ The reason for me to refrain from pinching you on the arm in order to prevent a bomb from exploding is obviously weaker than the reason for me to refrain from pinching you on the arm in order to annoy you. Preemption and Trolley may well be cases where an action causes harm, but because of the unusual circumstances, it is simply less important to avoid causing that harm. Thus, taking these additional considerations into account, it remains implausible to hold that in Preemption, Shooter does not harm Victim.

Fortunately, a proponent of the existence solution to the non-identity problem does not have to bite the bullet that Boonin bites in the case of Preemption. According to the existence account of harming, Shooter harms Victim: Victim has suffered a harm, and by shooting Victim, Shooter caused that harm. This is a major advantage of the view, and it counts towards thinking that the existence solution is more plausible, on the

15 See Gardner (2017).

whole, than Boonin's solution to the non-identity problem.

The other implausible implication of Boonin's view is illustrated by the following case:

Angry Alastair: Angry Alastair slaps you in the face. This is because he was trying to restrain himself; if he hadn't slapped you in the face, he would have torn your arms off.¹⁶

Recall, once more, that Boonin accepts P2, which tells us that A's act harms B only if A's act makes B worse off than B would otherwise have been. This principle implies that Alastair does not harm you by slapping you in the face, since if he hadn't slapped you in the face, you would have been even worse off. Boonin accepts this implication of his view and argues that it is more plausible that it might initially seem. His argument appeals to the following case:

Parking Meter. Your car is parked in a space with an expired parking meter. A police officer is about to write you a \$25 ticket, but then he notices that your door is unlocked and that there are four quarters in your cup holder. The police officer opens your door, takes the change, and feeds the meter. If he hadn't taken the dollar's worth of change from your cup holder, he would have written you the \$25 ticket.¹⁷

Boonin's intuition about the case is that the police officer doesn't harm you. But since Parking Meter is structurally similar to Angry Alastair, Alastair must not harm you in Angry Alastair, either.

Boonin's argument is not convincing. In the case of Parking Meter, even if we have the intuition that you aren't harmed, it's not clear that the best explanation for this intuition is that P2 is true. An alternative explanation is that losing a dollar simply isn't a big deal (at least for affluent people who own cars).¹⁸ Because it is easy to confuse a small harm with no harm, we need to test whether our intuitions are responding to the size of the loss or the absence of a necessary condition on harming. I propose that we

16 A slightly longer description of the case appears in Boonin (2014: 62).

17 This is my paraphrase of the parking meter case in Boonin (2014: 62).

18 An anonymous reviewer suggests a third explanation: that money is not the currency of harm. While I agree with the reviewer that the concept of "financial harm" is less compelling than the concept of harm to one's body or property, I disagree that Boonin's case relies on the assumption that money really is the currency of harm. Boonin elsewhere makes it clear that he thinks well-being is the currency of harm. Thus, in Parking Meter, I take him to be suggesting that when the police officer takes your dollar, your well-being drops by an amount that corresponds to the loss of one dollar. (And when I say "drops," I am making a temporal comparison, rather than a counterfactual comparison: your well-being is lower than it was before, but your well-being is not as low as it would have been had the police officer given you the ticket.) For more discussion about what we should take to be the currency of harm, see Tadros (2014).

consider a slightly different case:

Theft. Your car is parked in a space with an expired parking meter. Unfortunately, you have left a duffel bag containing \$5,000 in the backseat. A police officer is angry that about the expired meter, and he wants to ram his police car into your parked car, causing you \$8,000 in damage. However, he changes his mind and simply takes the duffel bag.

Here, you are out \$5,000, and I think it's clearer that you are harmed. Moreover, Angry Alastair more closely resembles Theft than it resembles Parking Meter, given the size of the harms at issue. Thus, it remains implausible to claim that in Angry Alastair, you are not harmed.

Here is how the existence account of harming supports the judgment that Alastair harms you: there is some general state of affairs either identical to or closely associated with the state of affairs in which you have been physically injured *at all*. If that state of affairs had not obtained, you would have been better off along some injury-related dimension of your well-being. Moreover, Alastair caused that general state of affairs, and so he harmed you. Importantly, the account does not necessarily imply that some *specific* upshot of the slap (such as a stinging cheek) is a harm to you. It may well be that if you hadn't suffered the harm associated with the stinging cheek, you would have been worse off along the same dimension on which the cheek pain registers.¹⁹ But in this case, it's reasonable to think that the slap is a benefit to you anyway, given that it prevents a greater harm.

In other words, in Angry Alastair, the harm you suffer is that your well-being is set back at all along a certain dimension, and not that your well-being is set back some specific number of units along that dimension. This implication of the existence solution is more plausible, I think, than the claim that you are not harmed in any sense.

6. CONCLUSION

Although Boonin critiques a number of different versions of the harm-based approach to the non-identity problem, he does not critique the existence solution. I have argued that the existence solution holds up well

¹⁹ On the other hand, if a stinging cheek corresponds to a completely different dimension of well-being than torn-off-arm pain, then there is a respect in which you would have been better off without the stinging cheek. In that case, Alastair's action causes you multiple harms: the general harm of being injured at all and the specific harm of the stinging cheek.

against the objections Boonin raises to the other harm-based views. I have also argued that the existence solution is more plausible than Boonin's own proposal for solving the non-identity problem. Nevertheless, the existence solution is committed to some controversial claims about counterfactuals and causation. To fully vindicate the existence solution, we will need a more complete account of the semantics of both foretracking and backtracking counterfactuals, and we will need to situate them within a plausible theory of causation that distinguishes between causes and mere conditions.

BIBLIOGRAPHY

- Adams, R., 1979: "Existence, Self-Interest, and the Problem of Evil", *Nous* 13: 53 – 65.
- Boonin, D., 2014: *The Non-Identity Problem and the Ethics of Future People*, Oxford: Oxford University Press.
- Brodbent, A., Unpublished book manuscript: *No Fire Without Smoke: Causation, Counterfactuals, and How to Make a Difference*.
- 2008: "The Difference Between Cause and Condition", *Proceedings of the Aristotelian Society* 108: 355 – 364.
- Elga, A., 2001: "Statistical Mechanics and the Asymmetry of Counterfactual Dependence", *Philosophy of Science* 68: S313 – S324.
- Gardner, M., 2015: "A Harm-Based Solution to the Non-Identity Problem", *Ergo* 2: 427 – 44.
- 2017: "On the Strength of the Reason Against Harming", *Journal of Moral Philosophy* 14: 73 – 87.
- Forthcoming: "When Good things Happen to Harmed People", *Ethical Theory and Moral Practice*.
- Hanser, M., 2008: "The Metaphysics of Harm", *Philosophy and Phenomenological Research* 77: 421 – 50.
- 2009: "Harming and Procreating", in *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*, ed. M. Roberts and D. Wasserman, 179 – 99, New York: Springer.
- Harman, E., 2004: "Can We Harm and Benefit in Creating?" *Philosophical Perspectives* 18: 89 – 113.
- 2009: "Harming as Causing Harm", in *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*, ed. M. Roberts and D. Wasserman, 137 – 54, New York: Springer.
- Hart, H.L.A. and T. Honoré, 1985: *Causation in the Law*, 2nd ed. Oxford: Oxford University Press.
- Hitchcock, C., 2007: "Three Concepts of Causation", *Philosophy Compass* 2: 508 – 516.
- Lewis, D., 1973: "Causation", *Journal of Philosophy* 70: 556 – 67.
- McGrath, S., 2005: "Causation by Omission: A Dilemma", *Philosophical Studies* 123: 125 – 148.
- Parfit, D., 1976: "On Doing the Best for Our Children", in *Ethics and*

- Population*, ed. M. Bayles, 100 – 15, Cambridge, MA: Schenkman.
— 1984: *Reasons and Persons*, Oxford: Oxford University Press.
- Shiffrin, S., 1999: “Wrongful Life, Procreative Responsibility, and the Significance of Harm”, *Legal Theory* 5: 117 – 48.
- Schwartz, T., 1978: “Obligations to Posterity”, in *Obligations to Future Generations*, ed. R. Sikora and B. Barry, 3 – 13, Philadelphia: Temple University Press.
- Tadros, V., 2014: “What Might Have Been”, in *Philosophical Foundations of the Law of Torts*, ed. J. Oberdiek, 171 – 92, Oxford: Oxford University Press.
- Thomson, J. J., 2011: “More on the Metaphysics of Harm”, *Philosophy and Phenomenological Research* 82: 436 – 58.
- Vihvelin, K., 1994: “Causes, Effects, and Counterfactual Dependence”, *Australasian Journal of Philosophy* 73: 560 – 573.