Modal Ethics

Part I: The Nonidentity Problem

Melinda A. Roberts

robertsm@tcnj.edu

Chapter 1

Intuition and Identity

1.1 *Goals, organization*. The most challenging nonidentity cases are those in which the act under evaluation is *clearly wrong* and the world in which that act is performed is *clearly worse* but we seem unable to find *any basis* on which to say *why* that's so *without* abandoning a certain deeply held intuition—the *person-affecting intuition.*

I believe that to solve the nonidentity problem but retain the person-affecting intuition we need to do two things. First, we need to recognize certain details—*modal* details—inherent in that class of most challenging cases, details about worlds *beyond* the world as it is and the world as it would otherwise have been. And, second, we need to formulate the person-affecting intuition, in both its deontic (act-evaluating) form and its telic (world-evaluating) form, by reference to principles that take exactly those modal details into account in determining the permissibility of the acts under scrutiny and completing our pairwise comparisons between worlds.

This Part I tries to accomplish both tasks. First, we situate our nonidentity cases in a modally *enriched* framework rather than a modally *impoverished* framework. And, second, we formulate the intuition itself in modally *sensitive* rather than in modally *constricted* terms.

It may seem that a clear implication of the modal approach is that pairwise comparisons between worlds $w\alpha$ and $w\beta$ may be affected, indeed *reversed*, by

events transpiring at still a third world wγ. Many philosophers will consider that implication highly problematic—and it does sound like I am asking for the recognition of spooky action at a *very* great distance. But we can also put the question in less mysterious terms. It's just whether a modally sensitive formulation of the person-affecting intuition is ruled out-of-bounds by an *axiological constraint* we have no choice but to accept. A third goal, then, of this Part I is to argue that it's not.

Thus in Chapter 2 below I outline how the most challenging nonidentity cases, as well as the nonidentity problem itself, that is, the argument to inconsistency, are standardly presented. Chapter 3 argues that that standard presentation reflects one or the other or perhaps both of two possible mistakes. Possible Mistake A is the mistake of under-describing, or misunderstanding, the case. Possible Mistake B is the mistake of thinking that how we formulate the person-affecting intuition is ruled out the above-mentioned axiological constraint.

Since the question of whether philosophers have indeed under-described, or misunderstood, their own cases can be settled only by reference to actual cases, I turn in Chapter 4 to what I regard as among the most challenging of the nonidentity cases, that is, Kavka's *pleasure pill case*. We can surely agree that in that case the act is *clearly wrong* and the outcome *clearly worse*. But it's very hard to say *why* the act is wrong and the outcome worse. It might seem that ridding ourselves of the person-affecting intuition would yield a quick solution—that to reject the intuition is to solve the problem. Perhaps that's so (though arguably it isn't). In any case, my argument here will be that a modally enriched understanding of the pleasure pill case, alongside a modally sensitive formulation of the person-affecting intuition, puts us in a position to solve the problem *without* rejecting the intuition. We never get to the result that the act under scrutiny isn't wrong or that the outcome itself isn't worse.

Chapter 5 briefly discusses still another class of nonidentity cases—cases that plausibly avoid Possible Mistakes A and B. I argue, however, that such cases don't meet the *clearly wrong*, that is, *clearly worse*, standard. If a properly

formulated person-affecting intuition implies in those cases that the act under scrutiny is *permissible*, it would not be unreasonable to consider the matter closed.

1.2 *Terminology. Possible worlds (futures; outcomes); distributions; accessibility.* We shall suppose that a given history of a given world (such as the history of the uniquely *actual* world) may unfold in many different possible ways going forward. We will call the many different ways in which such a history of such a world might unfold going forward *possible futures*, *possible worlds* or simply *possible outcomes*.

A *world* is not simply a *distribution*. A *distribution* is a bare-boned description of a world that simply (a) identifies the people that do or will exist at that world and (b) displays the overall lifetime wellbeing levels of each such person. A single distribution may apply to many different possible worlds. But a single world—itself a plethora of detail—determines a unique distribution of wellbeing across a unique population.

We can just note that distributions don't include information as to *other* distributions that agents might bring about in a given case. Worlds, though, are different. The world where agents are able (that is, have the ability) to snap their fingers and eradicate cancer is a distinct world from the otherwise similar world in which agents *lack* that ability. If agents have that ability at a world w1—if, we'll say, a world w2 where agents snap their fingers and cure cancer is *accessible* relative to w1—and agents don't have that ability a world w1′, then w1 and w1′ are distinct.

This last point can be expressed in the form of the *accessibility axiom*.[1] Thus, if in a given case w2 is accessible relative to w1, then in every case w2 is accessible to w1. Agents can't both, in w1, have the ability to snap their fingers and cure cancer and, also in w1, *not* have that ability; worlds are (unlike

_____

[1] See part 3.6.4.

distributions) far more finely differentiated than that. "Change a fact" about a given world and you have in effect changed the world you are talking about.

Accessible futures are (among other things) possible futures that are not barred by the laws of nature.[2] Not all *possible* futures, relative to a given history of a given world, are accessible. Thus relative to our own world history—the history, that is, of the uniquely *actual* world—the immediate future in which we snap our fingers and eradicate cancer is *possible*. There's no *logical* or *conceptual* inconsistency in the thought that that particular thwarting of—or sea change in—the laws of nature is right around the corner.   But that doesn't mean that the future in which we snap our fingers and eradicate cancer is now *accessible* to us. As things in fact are and, we think, will remain, that particular ability is one we don't have and have no means of acquiring.

Wellbeing; *personal good*; *general good. Wellbeing* indicates how good a person's existence at a given outcome is *for that person*. I think of *wellbeing* as that which makes a person's life so precious to the person who lives. For purposes here, we'll have no need to decide whether wellbeing consists in pleasure, happiness, preference satisfaction, resources, capability or something else entirely. But we will need to say a couple of things about wellbeing to proceed. First, a person p's having *more wellbeing* in an outcome wα than p has in an outcome wβ means that wα is *better for p* than wβ is—better for p, that is, "from p's own point of view" (whether or not *recognized* as better for p *by p*). Moreover, a person p's having a *positive* wellbeing level at an outcome just means that p's existence at that outcome is worth having. (Sometimes we'll just say that that person is *happy*.) And p's having a *negative* wellbeing level (p's being *miserable*) just means that p's existence is *less* than worth having—that that existence constitutes a *wrongful life*

_____

[2] Whether accessible futures are futures not barred by the acts of other agents is controversial. We might say that futures that aren't accessible to a given agent may nonetheless be accessible to a group of agents and thus consider the future to have a sort of derivative accessibility in respect of the individual agent. What we say on this question has implications for collective action problems, which I won't try to resolve here.

and that, from p's own point of view, it would have been better never to have existed at all.

The concept of the *personal good*, as distinct from *wellbeing*, will not come into play until Part II. There, we turn to the question whether the existence of an additional happy person in a given outcome makes that outcome better. But I should go ahead and note that *wellbeing* and the *personal good* are two very different things.[3] *Wellbeing* indicates how good a person's existence at an outcome is *for that person*. In contrast, the *personal good* indicates how good a person's existence at an outcome is *for that outcome*. Personal good, in other words, indicates how much value a given person's existence at a given outcome *contributes* to the *overall* value—that is, the *total* good, or what Broome calls the *general* good—of that outcome.[4] We said before that, if p has more *wellbeing* in one outcome than p has in another, then the one outcome is better *for p* than the other is. We can now note that simply in virtue of the meaning of the terms, if p has more *personal good* in one outcome than in another, then (other things equal) the one outcome will be *generally better* than the other.

Recognizing a distinction between *wellbeing* and the *personal good* leaves room for the idea that a person may have a positive *wellbeing* level in a given outcome even though that person's existence in that outcome contributes nothing at all to the *general* good of that outcome. In such a case, *wellbeing* may be positive even though *personal good* is zero. This point will be especially important when we turn, in Part II below, to the question whether the existence of an additional happy person in a given outcome makes that outcome better.

---

[3] I use the term *personal good* as Broome does. Broome 2015. Elsewhere Broome calls the personal good *wellbeing*—hence, the meaning he assigns to that term is distinct from the meaning I am assigning to that term here. Broome 2004. As far as I can tell, Broome from 2004 on has no analog to my *wellbeing* even though he makes reference to something's being good *for a person* (mainly to distinguish that from what he does want to talk about, which is the personal good, that is, a thing's being good *for an outcome*).

[4] The terms *personal good* and *general good*, and the relation that is defined between then, come from Broome 2015.

Acts; choices. Often I will use these terms interchangeably; often any act that implements a given choice will have the same morally relevant features as any other act that implements that same choice. In any such case, I will go back and forth freely between the terms *act* and *choice*. However, on occasion we will need to distinguish between *choices* and *acts*. I will then reserve *choice* as the umbrella term and keep in mind that any one of many different possible highly particularized *acts* performed at many different possible worlds may serve to implement a given *choice*. The nature of the agent's *choice* is often decided by the agent prior to the agent's performance of an act that implements that choice. But the nature of the *act* that implements that choice—what the agent ends up *doing*, in all its specificity, at a given world—typically isn't decided (at least, can't be *known*) until performance is complete. (Compare the *choice* to get a cup of coffee and the particular *act* that will, at the actual world, implement that choice; the *choice* is oblivious to *precise* motion, duration, etc., while the *act's* very identity depends on just those features.)

People. I include as *people* many non-human animals (for example, many mammals, birds and reptiles) in addition to many human beings. But the term also excludes some human beings, for example, human bodies that are alive but whose cerebral cortex is non-functioning.[5] I take for granted that a person who is *never conscious* at a given world *never exists* at that world.[6]

---

[5] See Peter Singer. [*Animal Liberation*; *Practical Ethics*.] The term *person* thus includes many nonhuman animals and excludes many human beings. For purposes here, I assume consciousness to be a necessary but not a sufficient condition for a thing's being a person.

[6] As just noted, for purposes here, I assume consciousness to be a necessary but not a sufficient condition for a thing's being a person. I assume, moreover, that to survive as the *same* person from one time to another—for the person p at t1 to be numerically identical to the person q at t2—is for consciousness to be knitted together in some fashion or another by a transitive relation of psychological connectedness R. Moreover, I take it that a human or non-human embryo or fetus that hasn't experienced consciousness isn't a person; a human or non-human fetus that *has* experienced consciousness is in close proximity of, but isn't identical to, a person; and the person that may ultimately develop out of a human or non-human embryo or fetus doesn't come into existence until consciousness emerges. Thus: early abortion involves never bringing a person into existence to begin with whereas late abortion might (depending on facts about when consciousness emerges in humans) involve removing a person from existence. Relying on that point, I have argued elsewhere

that, while abortion is certainly a matter of killing a fetus, the early or early middle abortion isn't a matter of killing a person but rather of never bringing a person into existence to begin with.

Chapter 2: Standard Presentation and Resolution of
the Nonidentity Problem

2.1 *Standard presentation of nonidentity facts*

      Population ethics owes much to two basic problems, the *procreative asymmetry* and the *nonidentity problem*. In Part II, I discuss the procreative asymmetry in the context of what John Broome calls the *neutrality intuition*. My focus there will be whether the strongly held intuition that the existence of an additional happy person doesn't, other things equal, make an outcome better leads to inconsistency. The focus of this Part I is the nonidentity problem.

      The cases that give rise to the nonidentity problem vary wildly in their specifics. Standard presentations of the facts of those cases, however, track the following outline.

<div style="border:1px solid black; padding:1em;">

<p align="center">Standard Presentation of Nonidentity Facts (Schematic)</p>

Let w1 be a possible future, or outcome, or *world*—say, the *actual* world, that is, the world as it *actually* unfolds. Let a1 be an act the agent performs at w1. Let p be a child born seriously impaired at w1 as a result of the agent's performance of a1 at w1.

Let w2 be a distinct world, a world available (we'll say *accessible*) to the agent. Let a2 be an act the agent performs at w2 in place of a1 at w1. Let q be a child *nonidentical* to p born healthy at w2 as a result of the agent's performance of a2 at w2 in place of a1 at w1.

We stipulate the following counterfactual: had the agent *not* performed a1 in w1, he or she *would have performed* a2 in w2 and q *would have existed* in place of p.

We stipulate as well that w2 is better for q than w1 is for p (we'll say that p has less *wellbeing* in w1 than q has in w2). We also stipulate that, despite the fact that p's wellbeing at w1 is suppressed as a result of the impairment, p's life at w1 is clearly worth living (p's wellbeing in w1 is clearly in the positive range).

From the fact that p has a life clearly worth living in w1 and never exists in w2, we infer that it's *not* the case that w1 is *worse for* p than w2 is. We also stipulate that no one *other than p* who does or will exist in w1 is affected in any way by what the agent does. Hence we infer that it's *not* the case that w1 is *worse for anyone* who does or will exist in w1 than w2 is.

[END]

</div>

Or, in graph form, where bold face means the indicated person does or will exist in the indicated outcome, and italics paired with the asterisk means the indicated person never exists in indicated outcome:

Graph 2.1: Standard Nonidentity Facts (Schematic)		
wellbeing	w1 (including a1)	w2 (including a2)
+10		**q**
+8	**p**	
+0	*q**	*p**

The *problem* arises when we combine the facts of any particular nonidentity case with the *person-affecting intuition*.

2.2 *Standard formulation of the person-affecting intuition; implications*

Let's first focus on the *deontic*, or *act*-evaluating, component of that intuition. The rough idea is that a morally "bad" act performed in a given world must be "bad for," that is, make things *worse for*, at least some person who does or will exist in that world.[7] An act that is bad *only* for the person who never exists at all—in virtue of its leaving *that* person out of existence altogether—cannot be bad, or wrong, at all.

Now, this sensible intuition is often formulated—I would argue reformulated—as the *highly constricted* PAIA(c).[8]

---

[7] See Parfit 1987, p. 363.

[8] Philosophers who have launched the nonidentity problem on the basis of PAIA(c) or an extensionally equivalent principle or set of principles include Shamik Dasgupta (forthcoming), part 1 (combination of claims (2) and (3)). See also Boonin 2014, p. 3 (discussion of premise "P2") and p. 52ff. (Chap. 3). ("The Counterfactual Account is the commonsense account of harm." (Boonin, p. 52)) See also Mulgan 2006 p. 8.

I agree with Dasgupta that an act's being "bad for" a given person p—in, we should add, a "morally relevant sense" (Parfit 1987, p. 374)—involves that act's making things *worse for* p. Thus, in my view, Harman and Shiffrin are mistaken in thinking that an act's

PAIA(c): aα performed at wα is morally wrong *only if* there is at least some person p such that p does or will exist in wα and aα performed at wα is "bad for"—that is, makes things *worse for*, p—than things would have been for p had aα not been performed.

I will argue, later on, that PAIA(c) is *unduly* constricted (part 3.4.1 below).

Let's now consider the *telic*, or *outcome*-evaluating, component of the intuition. The intuition here is that a morally "worse" world must itself make things "worse for" at least some person who does or will exist at that world.[9] The telic component, just like the deontic component, is also typically formulated in terms that are *highly constricted*.[10] Thus:

---

being bad for p can be explained in *non-comparative* terms. It doesn't follow, however, that we must adopt Dasgupta's—or Boonin's—*counterfactual* account of when an act is bad for p. After all, as we shall see in what follows, for reasons entirely independent of the nonidentity problem, that account is highly problematic. Rather, we should adopt a *modally sensitive* comparative account, one that determines that an act is bad for a person, not on the basis of what otherwise *would* have happened but for the act, but rather on the basis of what *could* have happened.

[9] See Parfit 1987, p. 370.

[10] See, e.g., Holtug 2010, p. 158. Holtug—and many other philosophers—go even further: w1 is *better* than w2 only if there is a person who does or will exist in w1 and w1 is *better for* that person than w2. See Holtug 2010, p. 158. See also Fleurbaey and Voorhoeve 2015 [in Hirose and Reisner, eds.], p. 102. (I assume that, when the authors write a "social situation cannot be better than another if it is not better for someone," they mean, "better for someone *who does or will exist in that situation*.)

But the principle that one outcome is better than another *only if* there is a person x who does or will exist in the one outcome and the one outcome is better for x than the other outcome is itself subject to instant counterexample. Thus, in cases involving *wrongful life*, where the person's life is clearly less than worth living and we want to say that, for that person, it would have been better never to have existed at all, the outcome that *excludes* that person is the better outcome even though no person who does or will exist in that better outcome such that that outcome is *better for* that person.

See also Arrhenius 2015 [in Hirose and Reisner]. Arrhenius thus explores the principle that an "outcome A is better (worse) than B" only if "A is better (worse) than B for at least one individual in A or B" (p. 111). I take it that principle implies that A is better than B only if A is better for at least one individual *in A*. If so, that means that this principle too is subject to the instant counterexample of wrongful life.

PAIO(c):   wα is morally worse than wβ *only if* there is at least some person p such that p does or will exist in wα and wα is *worse for* p than wβ is.

Applied to our nonidentity facts as those facts are standardly presented, those principles tell us both that a1 is *permissible* and that it's *not* the case that w1 is *worse* than w2, that is, that w1 is *at least as good as* w2 is.

I will argue that—like PAIA(c)—PAIO(c) is *unduly* constricted (parts 3.4.3 and 3.5 below). A less constricted version of the principle seems to capture the person-affecting intuition just as well—but is *far* easier to defend.

2.3 *Completing the problem*

So what's the problem? The problem is that we are *quite* confident that many cases that perfectly track the standard presentation involve acts that are clearly *wrong* and worlds that are clearly *worse than* the worlds we are comparing them against.[11]

Thus we agree with Parfit that the choices of depletion and the risky policy are wrong and with Kavka that it's wrong to sell your own future child into slavery or take the teratogenic pleasure pill prior to conceiving a child—even in the case where *refraining* from performing any of those wrong acts means the child whose plight we purport to be concerned about would never have existed—at least, *very probably* would never have existed—at all.[12]

Ditto for our pairwise comparisons of one world against another. We fully accept that the world in which depletion is implemented is morally worse than the world in which conservation is implemented, and that the world where the risky policy is implemented is morally worse than the world where the safe policy is implemented. And we accept that the world where the parent takes the pleasure

---

[11] David Boonin disagrees. See Boonin 2014 for the argument that cases that track the standard presentation involve acts that are, after all, perfectly permissible. See also David Heyd 2009 and Heyd 1992 for a metaphysically sophisticated argument reaching the same result.

[12] Parfit 1987, pp. 351-379; Kavka 1982, pp. 93-112.

pill and the one child is born impaired is worse than the world where the parent doesn't take the pleasure pill and a *nonidentical* but better off child is born instead.

Thus we face an inconsistency. We have reasoned our way to the results that a1 *isn't* wrong and that w1 *isn't* worse than w2 is. But we at the time fully accept that a1 *is* wrong and that w1 *is* worse than w2 is.

On the deontic side, the argument to inconsistency can be summed up as follows.

Standard Nonidentity Argument/Deontic Form		
Line no.		*Justification*
1	a1 performed at w1 is morally wrong.	Intuition
2	It's not the case that the act a2 that would have been performed at the world w2 had a1 not been performed at w1 makes things worse for p (or anyone else who does or will exist at w1) than a1 performed at w1 does.	Stipulations (life worth living; counterfactual; and no one else affected)
3	aα performed at wα is morally wrong *only if* there is at least some person p such that p does or will exist in wα and aα performed at wα is "bad for," that is, makes things *worse for*, p, than things would have been had aα not been performed.	PAIA(c)
4	It's not the case that a1 performed at w1 is wrong.	Lines 2 and 3
5	a1 at w1 both is and isn't wrong	Lines 1 and 4

To solve the problem is, in part, to avoid the inconsistency. But how to avoid the inconsistency?

The nonidentity argument in its telic form proceeds in parallel and similarly ends in inconsistency. Our moral instincts tell us that w1 *surely is* worse than w2— that's line (1). But PAIO(c) applied to the standard presentation of the facts instructs that w1 *isn't* worse than w2—that's line (4).

2.4 *Relation between deontic and telic nonidentity arguments*

I consider the nonidentity problem a problem for *both* the deontic *and* the telic components of the person-affecting intuition.

That way of thinking about the nonidentity problem doesn't seem to be a foregone conclusion. Thus many philosophers focus exclusively on the deontic argument. Those philosophers either (i) may consider the nonidentity problem to directly challenge *just* the deontic component of the person-affecting intuition or (ii) may not consider the project of ranking worlds in terms of their overall betterness a part of moral philosophy (they may, that is, be non-consequentialists). Philosophers subscribing to (i) may consider the telic intuition vulnerable to still *other* challenges but not clearly to the nonidentity challenge. But those same philosophers may instead leave the question of the telic intuition open even as they conclude that the nonidentity problem disproves the deontic intuition outright.

I believe, however, that the two discussions can't effectively be separated. Thus, as we shall see, the solution to the deontic form of the nonidentity problem that I will propose in what follows *requires* that we also, in effect, solve the telic form of the nonidentity problem as well.[13]

2.5 *Standard solution to the nonidentity problem*

The standard solution to the nonidentity problem has been to trace the inconsistency back to a strongly held intuition—that is, the person-affecting intuition—and then to reject that intuition.

The standard solution in the case of the procreative asymmetry has been to do the same—to trace the inconsistency back to the happy child half of the asymmetry; that is, the intuition that (other things equal) the existence of an additional happy child doesn't make the world better—and then to reject that intuition. That intuition itself is an implication of the *neutrality intuition*, which states that the existence of an additional happy person is morally *neutral*. Broome, correctly, argues that the neutrality intuition (understood as the neutral *range* claim)

---

[13] This argument is made in Chapter 5 below.

is false. He then concludes that that result (somehow) means that we should question—or outright reject—the happy child half of the asymmetry as well.

In Part II, however, I will argue that we can retain the happy child half of the asymmetry while avoiding Broome's objection by replacing the neutrality intuition with what I called the intuition of *narrow neutrality*. I will argue that narrow neutrality—the intuition *behind* the intuition; the intuition that the existence of the additional happy person does not make the world *better*, leaving open that such an addition may easily make the world *worse*[14]—is the only version of the intuition we have any interest in defending to begin with and that it's indeed a version of the intuition we can successfully defend. Through an inversion of Broome's concept of the *personal good*, we can even show that the intuition itself complies nicely with a certain constraint that Broome himself finds compelling, that is, Harsanyi's theorem.

The defense of the person-affecting intuition I propose in this Part I proceeds along similar lines. We can agree that highly constricted formulations of the intuition—such as PAIA(c) and PAIO(c)—fail. But those principles never accurately reflected the intuition to begin with. The problem with them is that they allow us to evaluate the case without ever attending to modal details inherent in those cases—details that may themselves have been, whether recognized or not, at the very root of our unwavering confidence that the act under scrutiny is clearly wrong and the one world clearly worse than the other. In contrast, by attending to the critical modal details of our own cases, and by formulating the intuition itself in modally sensitive rather than in modally constricted terms, we may give ourselves some chance of solving the nonidentity problem *without* abandoning some of what we were pretty sure we knew, that is, the person-affecting intuition itself.

---

[14] Thus narrow neutrality gives us the room we need to endorse the *second* half of the asymmetry—the intuition that the existence of the *miserable* child (other things equal) *does* make an outcome *worse*.

2.6 *Shared axiology*

Not coincidentally, the intuitions philosophers are quick to put on the chopping block—the person-affecting intuition and the happy child half of the asymmetry—are closely related. Both describe a necessary condition on when an act is wrong or an outcome worse: things must be made *worse for* some *existing or future* person.[15] The necessary condition is left unsatisfied in the case where the happy person in w2 happens to be an *additional* person (in what Parfit calls a "different number" case) and in the case where the *really* happy person q who takes the place of the *merely* happy person p happens to be *distinct* from q (in a "different people" case). With no one else there to satisfy the necessary condition, that condition is then failed. And we then obtain the problem results that the act we are confident is wrong is permissible and the world we are confident is worse is at least as good as the other.

The intuitions seem to have in common an underlying axiology. They both take the value of having more happiness than a world otherwise would have had but having that happiness stuffed into the container of an *additional* person or a *distinct* person—either way, a person q *nonidentical* to a person p—and discount that value to zero.

When, as a result of the nonidentity problem, philosophers reject that particular sensibility, they may consider the way cleared for the *impersonal* view that that value *cannot* properly be discounted: that *more happiness matters morally*, *regardless* of whether it's stuffed into the container of a nonidentical person or not.

---

[15] The person-affecting intuition and the happy child half of the asymmetry have much in common. In my view, however, *assuming that the intuitions are suitably formulated*, what they *don't* imply or have in common is the thought that, somehow, it's the people who do or will exist who matter morally, and not the people who will never exist at all. The view that some people matter morally and others not at all, whether articulated in the form of *strong* or *weak moral actualism* (see Hare 2007), does not stand up to scrutiny. The better view is that some diminutions in wellbeing—some *losses*, we might say—matter morally, while others matter not at all. For discussion, see Roberts 2010, Roberts 2011(a) and Roberts 2011(b).

That impersonal approach then is often translated into an *aggregative* approach—into the view that more happiness *in the aggregate* matters morally.[16] Some philosophers—including Temkin—fully accept that *other* values—e.g., equality—matter as well. But the lynchpin to solving the nonidentity problem, in Temkin's view, is to include the maximization of happiness *in the aggregate* as a value. That's the value we just can't get away from (in their view).

But of course it's *widely* understood that theories that abandon the person-affecting intuition—theories that insist on wrongdoing when all the agent has done is *not* bring an additional happy person into existence; theories that insist on wrongdoing when all the agent has done is *not* increase wellbeing on an *aggregate* basis—come with their own deficiencies.

Thus consider *totalism*, the paradigm example of the impersonal, aggregative approach. Many philosophers *like* the account totalism provides of the nonidentity problem.[17] (Totalists ask, "What problem?") But totalism is *widely*

---

[16] Nils Holtug is an exception. Holtug 2010. Holtug shows that we do not need aggregation to solve the nonidentity problem. See Holtug 2010 chaps. 6 and 9.

Holtug describes his approach person-affecting in nature. However, his view—as he points out—takes a wide, rather than a narrow, person-affecting form. He thus abandons the importance of *identity* to moral evaluation. According to Holtug, one outcome's being worse than another isn't necessarily a matter of that one outcome's being *worse for* any *particular* person. Rather, it's a matter of *either* the one outcome's being worse for one person *or* the other outcome's being better for *a possible distinct* person. Holtug 2010, p. 160.

When it abandons the identity condition, Holtug's view abandons the person-affecting intuition in what I take to be its most interesting form. It's that form of the intuition that I want to defend here against the nonidentity problem.

Fleurbaey and Voorhoeve (2015 [in Hirose and Reisner, eds.], pp. 103-105) similarly avoid aggregation but also abandon the identity condition.

[17] Not all philosophers think totalism does a good job with the nonidentity problem. Thus Elizabeth Harman argues that any satisfactory solution to the nonidentity problem will not simply generate the result that the act under scrutiny is wrong but also will provide an explanation of the wrongness of that act that is rooted in *what has been done to the child* we intuitively consider the victim of that wrong act. See Harman 2004, p. 90.

Holtug 2010 and Dasgupta (forthcoming) seem to agree with Harman that an adequate explanation needs somehow to reference the child whose plight has triggered our

understood to leave us fumbling badly when we consider what it has to say about some of the *many other* problems that arise in population ethics. Those include the *repugnant conclusion*, the *replaceability* problem and the *infinite population* problem. [18]

Sophisticated extensions of totalism do not do much better. Critical level theories must worry about the *sadistic conclusion*. And hybrid, or pluralistic, theories have a hard time generating any clear results at all and thus a hard time gaining our support. How can we accept a theory we *can't test*?[19]

---

concern. At the same time, though, on their views it's not what has been done to *that child alone* that accounts for the wrongdoing. Rather, it's also, in part, the fact that the one child is worse off than an alternate, better off child would have been.

[18] To see how difficulties with totalism—and averagism as well—may have motivated interest in developing an alternate approach, one that focuses on the "happiness of individuals rather than happiness on the whole," see Lazari-Radek and Singer 2014, pp. 363-73.

See also Parfit 1987, pp. 381-390. Most philosophers are completely nonplussed by the accounts totalism provides of the repugnant conclusion, replaceability and the infinite population problem. But some philosophers, I should note, bite the bullet and simply accept even what seem to be totalism's most unfortunate results. Thus, for an argument that we should accept the repugnant conclusion, see e.g. Huemer 2008.

[19] Larry Temkin takes the impersonal approach as far as he can in solving the problems of population ethics. Thus he describes what he calls the *internal aspects* view which recognizes the maximization of wellbeing in the aggregate (let's call this *additive maximization*) as a value but recognizes myriad other (impersonal) values as well (*equality*; *human flourishing*). Additive maximization means that the internal aspects view can offer a straightforward solution to the person-affecting person. But in the end Temkin makes a compelling case that the *essentially comparative* view—which itself recognizes a certain formulation of the person-affecting intuition—cannot be entirely set aside. *Both* approaches are, Temkin argues, in the end going to be called on if we are ever to solve the full range of problems in population ethics.

Hence what I call Temkin's *radical pluralism*. The testing challenges Temkin's theory faces are, however, profound. Not only does application of theory require that we balance many different *impersonal* values against each other (additive maximization against equality against human flourishing, etc.). It also requires that we balance all those *impersonal* values against *person-affecting* values. Temkin himself would be the first to acknowledge that the contemplated balancing procedure is itself going to be difficult to define and hence to test. See *Rethinking the Good* (Oxford 2012).

Both intuitions—both the person-affecting intuition and narrow neutrality—may thus have a critical role to play in moral evaluation. It thus makes sense to look hard at the arguments that purport to show that those intuitions fail before we cast them aside.

Depending on how we understand Temkin's overall argument, another question arises. If a theory X implies a world wα is at least as good as a world wβ is, and everything in X is true, then as a matter of logic X in combination with a theory Y also implies that wα is at least as good as wβ is. If knowing *less* supports a give result in a valid scheme, then knowing *more* supports exactly that same result. That means that the argument that (1) just focusing just on wα and wβ leads us to conclude, given X, that wα is at least as good as wβ is, but (2) then taking wγ in account leads us to conclude, now given X *and Y*, that wα *isn't* at least as good as wβ is (that wα is worse than wβ is), can't work. We need to reject (1) and say instead that we didn't validly conclude, given X, that wα is at least as good as wβ is; we need to say instead that we need to look beyond wα and wβ to complete the comparison.

If that is indeed Temkin's overall argument, then the right conclusion would not be that the internal aspects view and the essentially comparative view are both true and represent values that need to be weighed against each other, but rather than the internal aspects view is false.

Chapter 3

Critique of Nonidentity Argument

3.1 *Two possible mistakes*

I believe that the nonidentity argument to the result that a1 is both permissible and wrong fails. Ditto the argument to the result w1 both is and isn't worse than w2. And I believe that the widespread perception that the arguments succeed is rooted in one or the other or perhaps both of two serious mistakes.

The first possible mistake has to do with how philosophers have presented their own cases.

> *Possible Mistake A.* Philosophers have failed, starting out, to present the facts of their nonidentity cases—Depletion; Risky Policy; Slave Child; Pleasure Pill and many others—at an appropriate level of detail. In situating their cases within a *modally impoverished* framework—a framework that obscures or leaves unmentioned critical *modal details* inherent in their cases—rather than a *modally enriched* framework, they may have, in effect, under-described or perhaps even misunderstood the nature of their own cases.

Possible Mistake A creates two sorts of risks. First, the modal details that are left out of the under-described case might be details that are *critical* to solving the problem—*critical*, in particular, to avoiding the inconsistency *without* abandoning the person-affecting intuition.

Second, when critical details are left unrecognized, philosophers may be misled into thinking that still *other* details of the case *must* be *terribly telling* when in fact they are red herrings and tell us nothing at all. We basically start making things up when we give ourselves too little to work with starting out. We then design solutions that make *non*-critical details their centerpiece and thereby confine our solutions to ones that *can never actually work*.

The second possibility is this.

> *Possible Mistake B.* Philosophers have formulated the person-affecting intuition in terms of *highly constricted* principles—principles like PAIA(c) and PAIO(c); principles that, as we shall see, make the above-mentioned critical modal details irrelevant to the moral evaluation. I take for granted that philosophers have had some reason for going down that path. Perhaps, for example, they

think that the person-affecting theorist is bound by an *axiological constraint* that immediately rules out-of-bounds *modally sensitive* formulations of the intuition. But that assumption would itself be a mistake if, e.g., that axiological constraint doesn't have the reach or strength it's assumed to have.

If *that's* what's going on, then it really doesn't matter whether philosophers carelessly situate their cases within *modally impoverished* frameworks or conscientiously within *modally enriched* frameworks. Either way, they will consider the *axiological constraint* to bar any account that would bring those modal details to bear in solving the problem.

According, then, to Possible Mistake A, philosophers mistakenly think that *they* haven't arbitrarily limited the range of principles available to them for purposes of solving the nonidentity problem but rather that that range *just is limited* by the facts of their own cases. According to Possible Mistake B, philosophers mistakenly think that there's no point carefully ferreting out the modal details of their own cases since an axiological constraint will in any event make those details irrelevant for purposes of the moral evaluation.

I won't try to figure out here which possible mistake—A or B or both; missing facts or bad principles or both—is more probable. Rather, my purpose here is to show that avoiding both mistakes can help us make progress in solving the nonidentity problem *without* abandoning the person-affecting intuition.

3.2 *Possible Mistake A: Missing facts*

We might make the mistake of under-describing our own case. The standard presentation of the nonidentity case—Graph 2.1—makes just that mistake when it fails to specify whether the accessible outcomes for the agents are exhausted by w1 and w2. Consistent with that presentation, there may well exist some third accessible outcome w3 such that p is better off in w3 than p is in w1. If there is such a better-for-p w3, then it will also be true that, while w1 *isn't* worse for p than *w2* is, w1 *is* worse for p than *w3* is—and thus worse for p than *some other accessible outcome* is.

Now, not *all* cases that are thought to give rise to the nonidentity problem take that form. I believe, however, that the cases that give rise to the *most challenging* form of the nonidentity problem—cases, that is, in which the act under evaluation is *clearly wrong* but it's very hard to say *why*—take *exactly* that form. They are, that is, not *two outcome* cases, but rather *three outcome* cases.

That claim, of course, requires an argument. But it's not an argument that has any persuasive force made in a vacuum. We'll need to examine an actual nonidentity case. And that we shall do when we turn to Kavka's pleasure pill case (Chapter 4 below).

But in the meantime it's important to note that to accept that claim—to accept that the most challenging forms of the nonidentity problem are based on *three outcome* cases—is *not* to solve the nonidentity problem. Indeed, a number of philosophers accept, or seem at least prepared to accept, that their own cases include a better-for-p third accessible outcome w3.[20] They just think that the bare fact that w3 is *accessible*—that it's not at odds with the laws of nature; that it's something agent have some remote chance of bringing about; that it's *technically* accessible—doesn't take us very far at all in solving the nonidentity problem. I agree. The nonidentity problem is far more interesting than that.

3.3 *Setting aside third accessible outcome as irrelevant*

Thus philosophers who seem to accept a better-for-p third accessible outcome often nonetheless set that third outcome aside as irrelevant to the moral evaluation. Some philosophers offer one basis for that set-aside. Others at least suggest a second.

3.3.1 *Probability set-aside*. Philosophers sometimes offer *probability* as reason for analyzing a given nonidentity case as though it were a two outcome rather than a three outcome case. They may concede that some such better-for-p

---

[20] I include Kavka and Parfit here. Thus Kavka explicitly recognizes that agents might bring it about that one and the same child is better off, and Parfit asks us to accept that, after a few generations, depletion would in fact yield an entirely non-overlapping population but never asks us to accept that any such overlap is impossible.

w3 exists as an outcome that is *technically* accessible to agents. But they then push back: conceding w3 is accessible, they then argue that it's highly *improbable—so* improbable that for the purposes of analysis it can be ignored. Surely, even if agents *tried* to bring w3 or anything like w3 about—had agents, that is, *tried* to make things better for p than things are in w1—their chances of *success* would have been very close to zero. Very probably, they instead would have ended up leaving p out of existence altogether.

Kavka thus described the *precariousness* of the coming into existence of any particular person. Had agents done things otherwise than *just as they did* in the period of time leading up to the conception of the *particular* person p, they would thereby have affected—changed—at least in some slight way the *timing and manner* of conception.[21] And affecting the timing and manner of conception in even the slightest way would surely have reduced *p's* chances of coming into existence to almost nothing at all. (It's the males, mainly, who are behind the precariousness of existence, producing 200 million plus sperm cells per sexual encounter, a distinct inseminating sperm cell, it seems reasonable to suppose, yielding a distinct person.)

From that probability point, philosophers segue to the position that for all practical purposes, including moral evaluation, we might as well just ignore w3— that we may as well analyze the particular nonidentity case as a two outcome rather than a three outcome case.

But that segue isn't valid. We are familiar with and should accept the logic that (presumably) is *supposed* to support the inference. Consider the following medical example. If Ginger is sick and facing her own imminent demise, she prefers a treatment X that promises a very high probability of survival *and* a very good life if she *does* survive in place of a treatment Y that promises that, if she survives, her life would be *even better* than "very good" but also promises almost *no chance of survival at all*. Ginger, correctly, considers treatment X so clearly

---

[21] See Kavka 1982, p. 93; Parfit 1987, p. 361.

better for her that she doesn't think there's any *practical* reason at all for even keeping treatment Y on the table as an alternative worthy of any further serious consideration.[22]

We can and should accept that logic. The problem is that it doesn't apply to the most challenging versions of the nonidentity problem.

That the low probability of the agents bringing about the better-for-p *w3* means that we can set w3 aside as irrelevant for purposes of our evaluation is a mistake *unless* we *also* happen to know that the probability of the agents bringing about the not-quite-so-good-for-p *w1* is somehow *greater*.

Thus it's *stipulated* in the medical example that treatment X would give Ginger a *very high probability* of survival. But nothing in the most challenging versions of the nonidentity problem tells us that the probability of the agents bringing about w1—or any outcome in which p exists and has a very good life—is itself *very high*.

It might be argued that w1's high probability (versus w3's low probability) is just *presumed* as a fact of the case—a fact we are absolutely free to make perfectly explicit in the standard presentation if we happen to suffer from OCD and thus feel ourselves compelled to spell out such an obvious point.

But how can we *presume* that the precariousness of existence applies to p's coming into existence given *a3* but (somehow) *doesn't* apply to p's coming into existence given *a1*? Why should we think that the *wrong act* a1 will somehow make it more probable—measured at the appropriate moment; that is, the moment just prior to performance—that p will eventually come into existence than will at least *some* other act that we take to be *better* for p, e.g., a3?

If it's then correct in the particular case that the precariousness of existence cuts both ways—that the probability of p's coming into existence is very low whether a1 is performed or a3 is presumed; that probability of w1, given a1, or, more accurately, given a choice that *includes* a1, is just as low as the probability of

---

w3, given a3, or, again, a choice that *includes* a3[23]—then the fact that the better-for-p w3 exists as an accessible outcome becomes *highly significant*.

To see that that's so, we just need to slightly revise the medical example: if you are ill and the probability of cure, miniscule under Y, is equally miniscule under treatment X, then of course you prefer treatment Y. We all prefer (other things equal) the very small chance of winning the *billion* dollar lottery over the same very small chance of winning the *million* dollar lottery.

This point can be put in terms of *expected wellbeing*, where the *expected wellbeing* of a given act for a given person is just the summation, for each possible outcome of that act, of the actual wellbeing assigned to that person by that outcome multiplied by the probability that that outcome will obtain, given that act. Thus the *expected wellbeing* X produces for you in the second version of the medical example is greater than the *expected wellbeing* that Y produces for you. Ditto the lottery example. The *expected wellbeing* produced for me when I pay a dollar for a very small chance at a billion dollar jackpot is greater than the *expected wellbeing* produced for me when I pay the same dollar for the same very small chance at a million dollar jackpot.

This is not, of course, to argue that *all* future-directed acts share the identical, *very low* probability of bringing any particular person into existence. We can easily construct a case in which the probability of w1 *is* greater than the probability of w3—or, more precisely, where the probability of the agent achieving the lesser outcome for p is greater than the probability of the agent achieving the better outcome for p. And all we need, for purposes of showing that the person-affecting intuition is false, is *one* successful counterexample. Effective, but risky, fertility treatments—treatments that increase the chances of conception but also increase the chances of the child's being burdened in some way if he or she is

---

[23] More accurately: if the probability of p's coming into existence is very low whether the agent makes the choice that includes a1 or the choice that includes a3—if, that is, the probability of w1, given the choice that *includes* a1, is just as low as the probability of w3, given the choice that *includes* a3—then w3's accessibility becomes highly relevant to the analysis, notwithstanding its low probability of obtaining.

conceived—come to mind. Depending on the actual numbers, such treatments may well generate more expected wellbeing for a particular child than a lower-risk but largely ineffective alternative. But is it really so clear—provided the risk itself is minimized (that is, that there is no still better alternative for the child) and that there's no chance of anyone else being affected in any way—that it's wrong for the physician to provide or for the patient consent to the effective, high-risk treatment? I don't think so.

The upshot here is that we can't presume the probabilities one way or the other in advance of our close scrutiny of the details of the particular case. We certainly can't *presume* that the phenomenon of the precariousness of existence *only* applies when the act under scrutiny is one we consider *clearly permissible* and that it somehow disappears into the woodwork when the act is one we are confident is *clearly wrong*. Nor can we presume that the probabilities are always a wash. We shall instead have to look at the cases.

I think then what we will find—and that is the topic of Chapter 4—that the cases that pose the only *clear* challenges to the person-affecting intuition—the cases, that is, in which the choice *is* clearly wrong; and here I include depletion, risky policy, slave child, pleasure pill, historic injustice and many others—are cases in which the (very low) probability of the agent's achieving the not-quite-so-good-for-p w1 on behalf of p is *no greater than* the (very low) probability of the agent's achieving the better-for-p w3 on behalf of p.

3.3.2 *Counterfactual set-aside*. Other philosophers may set w3 aside as irrelevant for reasons having nothing to do with probability. They may think that w3 is irrelevant because they consider the following counterfactual highly relevant to the analysis: had agents refrained from performing a1 and thus failed to bring about w1, then the agents *would have performed* a2 at w2 instead. They may think, in other words, that in view of that counterfactual what's going on at *w2*—but *not* what's going on at *w3*—is relevant to the discussion.

Exactly that counterfactual is included as a stipulation in the standard presentation of the nonidentity case.

But it's a mistake to think that that counterfactual makes w3 irrelevant to the evaluation. That I *would* have done still worse to a given person p, had I not just as I did, is never a vindicator of what I have in fact done *unless* it's also true that I had no third, better-for-p alternative. When I do have such a third, better-for-p alternative, whatever I *would* have done, what I *have* done may well still be wrong.

Suppose, for example, that I shoot Harry in the arm and that, had I not shot him in the arm, I would have shot him in the heart. Surely, in that case, my shooting Harry in the arm is permissible *only if* I somehow don't have the third alternative of just standing there and not shooting Harry at all. It's the existence or non-existence of that third alternative that determines whether my shooting Harry in the arm is permissible—*not* the fact that I *would* not have availed myself of that alternative had it existed. (The *world* decides what is permissible, not what happens to be convenient for the *agent*.)

3.4 *Possible Mistake B: Bad principles*

Here, we must worry about whether the *deontic* and *telic* components of the person-affecting intuition are properly formulated.

3.4.1 *Bad deontic principle.* According to the deontic, or act-evaluating, component of the person-affecting intuition PAIA(c), aα performed at wα is morally wrong *only if* there is at least some person p such that p does or will exist in wα and aα performed at wα makes things *worse for p* than things *otherwise* would have been—than things, that is, would have been *but for* wα.

> PAIA(c):   aα performed at wα is morally wrong *only if* there is at least some person p such that p does or will exist in wα and aα performed at wα is "bad for"—that is, makes things *worse for*—p than things would have been had aα not been performed.[24]

By directing us just to compare what agents *have done* to a given person against what those agents *would have done* to that person had agents not done the thing they *have* done, the highly constricted PAIA(c) obviates any need to guard against Possible Mistake A. By its very nature, PAIA(c) makes any accessible outcome beyond wα and wβ irrelevant to the analysis.[25]

---

[24] For philosophers who have their discussions of the nonidentity problem with a formulation of the person-affecting intuition very like PAIA(c), see note 8 above. Parfit is an exception.

[25] There are, of course, other unfortunate formulations of the person-affecting intuition in addition to PAIA(c). One such formulation noted and then rejected by Lazari-Radek and Singer, and previously Singer, is the *prior existence view*. Lazari-Radek and Singer 2014, p. 368; Singer 2011, pp. 88-89. They argue, convincingly, that the miserable child half of the procreative asymmetry shows that the prior existence view will not work.

More generally, the miserable child case and many others convince us that *any* principle that tries to draw a distinction between people who have moral status and people who don't—on the prior existence view, the miserable child doesn't have moral status in virtue of the fact that that child isn't among those people who do or will exist *however* the choice under scrutiny is made—will fail. That would include views that deem only *actual* people (whether existing or future) to have moral status and views that deem only the

Philosophers who think that the person-affecting intuition is to be understood by reference to PAIA(c) then have little difficulty convincing us that PAIA(c) generates false results in many nonidentity cases.

The difficult with their argument is that there is no good reason to think that PAIA(c) captures the person-affecting intuition—that the necessary condition expressed by the person-affecting intuition is as stringent; that is, as easily failed— as PAIA(c) says that it is.

Thus, the person-affecting intuition is best understood *not* to imply that what the agent has done is permissible *whenever* what the agent *would have* done, had the agent not done just what the agent did, makes things *still worse* for the person. Rather, it's best understood to imply permissibility just when the agent had *no way at all* of making things *better* for that person.[26] After all, we can surely agree that PAIA(c) is simply false.

We needn't mine the area of population ethics in order to see why that is so. The shoot-Harry-in-the-arm case does that work for us. It's the fact that I have a third alternative in that case—the alternative of just standing there and not shooting Harry at all—that shows that my shooting Harry in the arm was wrong.

And there's a larger lesson here as well: we can't accurately evaluate my shooting Harry in the arm until we have *looked around at all the details of our own case*—including the *modal* details; the details regarding what *could* have been— and noted that I had the alternative of not shooting at all. Had I not had that

---

people who do or will *exist under the act under scrutiny* to have moral status. For more on *moral actualism*, see note [15] above.

But not all formulations of the person-affecting intuition make that mistake. PAIA* and PAIA** don't. They avoid the result that it's permissible to bring the miserable child into existence, and they at the same time imply that it's not wrong not to bring the happy child into existence.

Critically, however, those same principles *are* vulnerable to the *nonidentity problem*. Hence the motivation for this Part I.

[26] The person-affecting intuition is best understood, in other words, to open the door to a finding of wrongdoing whenever agents fail, for each person who does or will exist, to *maximize* wellbeing for that person.

alternative, my shooting Harry in the arm would have been permissible, indeed, obligatory.

If that point holds for the easy, ordinary, *same* person case, then why it would not also hold for the harder, extraordinary, *additional* person case is very unclear.

It seems, then, that we should formulate the deontic component of the person-affecting intuition by reference not to the *highly constricted* PAIA(c) but rather to the *modally sensitive* PAIA*.

> PAIA*: aα performed at wα is morally wrong *only if* there is at least some person p such that p does or will exist in wα and there is an alternative act aβ performed at an alternative accessible outcome wβ such that aα performed at wα makes things *worse for* p than aβ performed at wβ does.[27]

In other words: if aα performed at wα *maximizes* wellbeing for each person who does or will exist at wα, then aα *isn't* wrong.

PAIA* in hand, I can now outline the response to the nonidentity problem that I think works. For the most challenging forms of the problem—those in which the act under scrutiny is *clearly wrong*—we regularly find that we are perfectly able to identify a third, better-for-p alternative, an alternative beyond what the agent in fact does and beyond the counterfactual alternative the critic of the person-affecting intuition wants to restrict our attention to. We then simply note that, in such a case, PAIA* avoids the implication of permissibility and thus opens the door for a finding of wrongdoing based on still other, widely accepted, person-affecting principles.[28]

---

[27] I am happy to say that, if there is such an act aβ at wβ, that aα *harms*, or imposes a *loss* on, the person it makes things worse for. It would be a mistake, however, to get bogged down in the meaning of the term *harm* or *loss* and hence my use of those terms may be regarded as simply shorthand for making things *worse for* a given person in the sense described by PAIA*.

[28] Pareto principles are good examples here. Thus where w3 is accessible relative to w1, and w1 and w3 contain the same people, and w3 is better for at least one person and worse for none, we will say that the act the produces w1 is itself wrong. For discussion, see Roberts 2010 (*Abortion and the Moral Significance of Merely Possible Persons*).

3.4.2 *Objection based on probabilities*. It might seem that PAIA* fails to capture an important element of the person-affecting intuition. That PAIA* itself sidesteps a finding of permissibility in a given nonidentity case is not very useful if some other important element of the person-affecting intuition generates exactly that result.[29]

Specifically, it might seem that PAIA* is blind to the above-mentioned phenomenon of the *precariousness* of existence. It's one thing to require, for wrongdoing, that a better-for-p alternative exist as a *technically accessible* outcome. That requirement is—we might well agree, at least for the most challenging versions of the nonidentity problem[30]—relatively easy to satisfy. And indeed my argument—in Chapter 4 below—will be that for those versions of the problem the requirement *is* satisfied.

But surely the person-affecting intuition isn't just about *technical accessibility*. Surely it also has something to say in the case where the agent can *technically* bring about the better-for-p outcome but where, *whatever* the agent does, the *probability* that that better-for-p outcome will obtain is *very, very low*. In other words, surely the person-affecting intuition includes the idea that an act—say, a1 at w1—is wrong *only if* agents have *at least as good a chance* of bringing about the better-for-p w3 as they have of bringing about the not-quite-so-good-for-p w1. Roughly put, the idea here is this: only when the alternate, technically accessible, better outcome is an outcome agents had *some significant chance* of bringing about do we have room to declare that the one act leading to the one outcome is *wrong*.

---

[29] [where to place?] We noted earlier that the fact that a better-for-p accessible outcome is *highly improbable* doesn't mean that we should deem that outcome *irrelevant* for the purposes of moral evaluation. Rather, the legitimacy of setting aside any such improbable better-for-p outcome as irrelevant will depend on whether any not-*quite*-so-good-for-p alternative accessible outcome is any *less* improbable. I think we put that particular point to bed at least for purposes of *theory*. As a *theoretical* matter, that point is indisputable.

[30] There are plenty of other nonidentity cases in which it's failed. In my view, however, those cases are not among the most challenging. For further discussion of this other sort of nonidentity case, see Chapter 5 below.

Now, the *actual* value person-affecting theorist might resist that point, taking the position that PAIA* exhausts the deontic component of the person-affecting intuition. But many philosophers want their moral principles to be *action-guiding*. The question is whether a person-affecting complement to PAIA* can be formulated that can play that that action-guiding role *without* falling prey to the nonidentity problem.

What we need, then, is a principle—we'll call it *PAIA*** in what follows—that considers *expected* wellbeing—not just *actual* wellbeing—to have a role to play in determining permissibility.

But in designing PAIA** we should insist on one condition—the *level playing field* condition. We should insist that analysis under PAIA** *not* be weighted in *favor* of the act under scrutiny—that is, the wrong act—and *against* each alternative to that act. Thus, if we decide that probabilities are important, and if, on that basis, we decide that what is relevant to our moral evaluation of a1 is the *expected* wellbeing generated for p by *each alternative to a1*, then so must we be committed to the position that what is relevant to our moral evaluation of a1 is the *expected* wellbeing generated by *a1*.

Specifically, we should resist the thought—tempting as it may be, in a *post hoc ergo propter hoc* sort of way, given that how the future *actually* unfolds at w1 is *stipulated* as part of the case; so tempting that what we really face here might be called the nonidentity *fallacy*—that evaluating a1 is just a matter of comparing the *actual* wellbeing a1 generates for p at w1—very high, since p has a very good life in w1—against the *expected* wellbeing each *alternative* to a1 generates for p at each alternative accessible world—*very* low, given the very long odds against any one person ever being conceived at all *however* agents comport themselves. Rather, it's the *expected* wellbeing of a1 for p that should be compared against the *expected* wellbeing of each alternative to a1 for p.[31]

---

[31] For those philosophers who *don't* find it reasonable to think that probability has a role to play in determining permissible, we would be happy with the actual wellbeing against actual wellbeing comparison. The point here is that what won't do is mixing apples and oranges in this context.

Thus, according to this new principle, if aα performed at a world wα *maximizes expected wellbeing* for each person who does or will existing in wα, then aα *isn't* wrong.

PAIA** expresses that idea.

> PAIA**: aα performed at wα is morally wrong *only if* there is at least some person p such that p does or will exist in wα and there is some alternative act aβ performed at an alternative accessible world wβ such that the expected wellbeing of aα performed at wα for p is less than the expected wellbeing of aβ performed at wβ for p.[32]

Accepting PAIA** as an element of the person-affecting intuition raises the stakes. It means that for the most challenging forms of the nonidentity problem—those in which the act under scrutiny is *clearly* wrong—to defend the intuition I will need to show that we really can identify an alternative act performed at an alternative accessible world that satisfies the necessary condition on wrongdoing that PAIA** provides. I will need to show, that is, that we can identify an alternative act that generates still more *expected* wellbeing for the person whose plight has concerned us starting out than the act we agree is clearly wrong does. I turn to that work in Chapter 4 below. But first we need to consider the telic, or outcome-evaluating, component of the person-affecting intuition.

I have thus elsewhere argued that an approach that determines moral permissibility on the basis of a comparison between one act's *actual* value and another act's *expected* value is inconsistent. The point is an obvious one but widely disregarded in the nonidentity literature.

[32] For my own part, I would accept both PAIA* and PAIA**—accept, that is, that, for an act at a world to be wrong, there must be some person p who does or will exist at that world and some alternate act performed at some alternate accessible world such that both the actual and the expected value of the alternate act for p is greater than the actual and the expected value of the one act for p.

3.4.3 *Possible Mistake B: Bad telic principle*. Philosophers often formulate the telic, or outcome-evaluating, component of the person-affecting intuition in the form of PAIO(c).

> PAIO(c): wα is morally worse than wβ *only if* there is at least some person p such that p does or will exist in wα and w1α is *worse for* p than wβ is.[33]

But why should we think that the telic component of the intuition must be forced in the mold of the highly constricted PAIO(c)?

Why not, that is, immediately abandon PAIO(c) as, e.g., a typo and go straight to PAIO*:

> PAIO*: wα is worse than wβ, only if there is some person p who does or will exist in wα and some accessible outcome wγ (which may but need not be identical to wβ) such that wα is worse for p than wγ is.

In connection with the evaluation of *acts*, the person-affecting intuition is best understood to require us to look around at *all* the facts of our own cases before deciding permissibility rather than focusing exclusively on what *was* done and what otherwise *would have been* done. Why not say the same thing about the evaluation of *outcomes*—that there, too, the person-affecting intuition is best understood to require us to look around at *all* the facts of our own cases before deciding whether any one outcome is morally worse than any other?

To my knowledge, critics of the person-affecting intuition don't explicitly answer (or raise) that question. Rather, they simply formulate the telic component in terms of PAIO(c) or a similarly highly constricted principle and then proceed to the (rather straightforward) task of refuting that principle.[34]

---

[33] For philosophers who have proposed this formulation, see note 10 above.

[34] Thus I think we can, and should, reject PAIO(c). But my main goal is to show that we can reject PAIO(c)—which was never an adequate way of capturing the person-affecting

We are thus left to speculate. My own sense is that the answer to the question has little to do with the nonidentity problem or the person-affecting intuition and everything to do with a certain *axiological constraint* that many philosophers, including many critics of the person-affecting intuition, consider applicable to any proper pairwise comparison between outcomes. They, in other words, assume that PAIO* is ruled out-of-bounds by an *axiological constraint* we have no choice to accept. I consider that possibility now.

3.5 *Axiological constraint*

3.5.1 *Source of constraint; the classic view.* Since the distinction between PAIO(c) and PAIO* becomes apparent only in the context of three outcome cases—they generate the same result in any *two* outcome case—I focus on the three outcome case here.

---

intuition to begin with—without rejecting the intuition. We shall simply do a better job articulating the intuition. We should prefer PAIO* to PAIO(c).

Graph 3.5.1: Three Outcome Schematic			
wellbeing	w1 (including a1)	w2 (including a2)	w3 (including a3)
+10			**p**
+5	**p**		
+0		*p**	

It's easy enough to produce an argument that concludes with the axiological constraint. We can do it in two quick steps.

The *first step* is to accept what I will call the *classic view* regarding how any proper pairwise comparison between outcomes is to proceed. The classic view is reflected in totalism. But it's a view that many maximizing consequentialists—including many pluralists—share.

Maximizing consequentialists are of course accustomed to what I would consider *modally sensitive principles* when it comes to the evaluation of *acts*. No one thinks that the fact that aα is better than aβ means that aα is *permissible*.

Things change, however, when we turn to the evaluation of *outcomes*. According to the classic view, the pairwise comparison between accessible outcomes can be completed—not *must* be completed; simply that, in contrast to the case of *acts*, there is no reason the pairwise comparison *can't* be completed since there the goal is not to determine which outcome is *best* but rather to determine which outcome is *better*—by examination of *just* the *two* outcomes and *without* reference to any *third* outcome.

But of course, if the comparison between wα and wβ *can* be completed without examining any third accessible outcome wγ or indeed even knowing whether such a third wγ exists, then that would mean that we don't *need* to look beyond wα and wβ to rank wα against wβ. But if we don't *need* to do that, then

anything beyond wα and wβ *cannot make a difference* to how we compare wα against wβ. And if nothing beyond wα and wβ *can* make a difference to how we compare wα against wβ, then allowing considerations relating to wγ—those *modal details*—to affect—that is, to *change*—how we rank wα against wβ would be irrational.

And it's that last point—that core component of the classic view—that I want to focus on here: allowing considerations relating to wγ—those *modal details*—to affect—that is, to *change*—how we rank wα against wβ would be irrational.

The *second step* is then to squeeze the person-affecting intuition into the classic view—that is, to give the person-affecting intuition the appropriately person-affecting role to play a role in ranking w1 against w2 *but* at the same time making sure that that ranking can be completed *without looking beyond w1 and w2*.

The result is a commitment to the position that that the person-affecting necessary condition on when w1 is worse than w2 can be satisfied *only in a certain way*. Specifically, it's a commitment to the following position:

> *Axiological constraint.* The person-affecting necessary condition on w1's being worse than w2 can be satisfied *only if* w1 is worse for a person p who does or will exist in w1 *than w2 is*.

Thus the *axiological constraint*—a restriction on how the telic component of the person-affecting intuition is to be formulated.

PAIO(c), of course, satisfies that constraint. But it's a constraint that PAIO*—which explicitly takes what is going on at w3 into account in ranking w1 against w2—immediately fails.

Now, we'll need to examine the axiological constraint. Is it really one that the person-affecting theorist has no choice but to accept? We will turn to the arguments in part 3.6 below.

But first we attend to some miscellaneous matters: underlining why the axiological constraint represents such a deep challenge against a person-affecting approach (3.5.2); noting (relatedly) that the options for maintaining a person-affecting approach to the nonidentity problem without discarding the axiological

constraint are unpalatable (3.5.3); and finally articulating the implications the axiological constraint has for the particular version of the neutrality intuition—that is, *narrow neutrality*—I defend in Part II.

3.5.2 *How the axiological constraint undermines a person-affecting solution to the nonidentity problem.* Under plausible assumptions, accepting the constraint means that our person-affecting deontic solution to the nonidentity problem—outlined in Chapter 2 above and filled out in Chapter 4 below—cannot be made to work. Those assumptions include (i) that there exists a *very tight connection* between act evaluation and outcome evaluation; and (ii) that the *at least as good as* relation between outcomes is *transitive*.

Why? The very tight connection between act and outcome evaluation makes the following procedure possible and indeed natural. We first rank the outcomes. We then determine among those outcomes which are accessible. And finally we evaluate the acts that bring those outcomes about. Thus an act that produces a given outcome is wrong *if and only if* that outcome is worse than at least some other accessible outcome is.[35] Applying that procedure to a three outcome case—Graph 3.5.1—we understand that, if $w1 < w3$ and $w3$ is itself accessible, then $a1$ is wrong. Now, while the *procedure* is to rank the outcomes and then evaluate the acts, the *inferences* go in both directions. Thus, if $a1$ at $w1$ is wrong, then $w1$ itself must be worse than at least *some* other accessible outcome.

We then compare $w1$ against $w2$. According to PAIO(c), it's not the case that $w1$ is worse than $w2$, there being no existing or future person p in $w2$ such that $w2$ is worse for p than $w1$ is. Hence: $w1$ is at least as good as $w2$; $w1 \geq w2$. But

---

[35] Some philosophers who question the very tight connection between act and outcome evaluation do so on the basis of the special obligations they think we have in respect of our own nearest and dearest. But I find *impartiality* the more plausible view. See Hirose, "A Puzzle from Nagel's Pairwise Comparison" [in AAA Research]. And I see no reason that we cannot *both* accept impartiality *and* a well-articulated, modally sensitive person-affecting approach—accept, that is, impartiality *but also* accept that identity is important. See note 1 above.

PAIO(c) also instructs both that $w2 \geq w1$ and that $w2 \geq w3$, there being no existing or future person in $w2$ for whom $w1$ or $w3$ is worse than $w2$ is.

Assuming transitivity, we then infer that $w1 \geq w3$.

The problem for the deontic person-affecting solution to the nonidentity problem may now be obvious. For that solution to work, a1 at w1 can't be—and wasn't, under PAIA*—deemed permissible. Avoiding the implication of permissibility opens the door to still other person-affecting principles that would evaluate a1 at w1 as *wrong*.[36] But if a1 at w1 is wrong, then the *tight connection* between act and outcome evaluation means that w1 must be *worse* than at least *some* accessible outcome. Here the obvious candidate is w3 (though an analogous problem arises if we go with w2 instead); hence $w1 < w3$ ($w1 < w2$). But w1 can't both be at least as good as w3 is *and* worse than w3 is. (w1 can't both be at least as good as w2 is *and* worse than w2 is.)

In effect, the move to PAIO(c) completely dismantles the person-affecting deontic solution to the nonidentity problem. Not only does it bar what it seems the person-affecting theorist should want to say about the *outcomes* that are to be compared. It also means we must, after all, endorse the result that the *act* a1 is permissible—that is, reject the very result that we earlier said that PAIA* leaves room for: that a1 is wrong.[37]

---

[36] See note 28 above [Pareto principles].

[37] To see, in other terms, why accepting the axiological constraint disrupts our prior solution of the nonidentity problem, consider the following.

If w3 is the outcome that demonstrates that a1 is wrong, then $w3 > w1$.

But if our interest is in comparing w1 and w2, and if that means we can't know about w3, then we can't know that $w3 > w1$. Instead we will think that w1 is at least as good as w2 is. (We can't see anything amiss in in w1.)

I.e., we can't see the truth about w1 and w2; we can't see that $w1 < w2$, that is, that w1 is worse than w2 is.

But if w1 is deemed at least as good as w2 is, and if we have already deemed w2 permissible under our person-affecting theory, then w1 must be permissible as well. But

In contrast, PAIO* works with PAIA* hand in glove. The necessary condition on w1 being worse than *w2* that PAIO* sets forth is perfectly satisfied: there does or will exist a person p in w1 such that w1 is worse for p *than w3 is*. PAIA* thus opens the door for a further person-affecting principle that deems w1 worse than w2 on the ground that w1 is worse for p *than w3 is*. More generally, we can say just what we want to say about this case: w2 is exactly as good as w3, and w1 is worse than both w2 and w3.

3.5.3 *Unpalatable options for preserving the person-affecting solution to the nonidentity problem.* A theorist completely smitten with the axiological constraint but finding the person-affecting intuition attractive and hence wanting to articulate that intuition carefully might consider one or both of the following options attractive.

(i) One way of preserving our solution to the nonidentity problem while accepting PAIO(c) in place of PAIO* would be to reject the assumption of *transitivity*. Indeed, it might seem that the preceding discussion has shown that that's just what the person-affecting theorist *should* so in order to preserve the person-affecting solution to the nonidentity problem. It might even seem that the preceding discussion shows that to accept the person-affecting approach just is to reject transitivity.[38]

---

that's an inconsistency, since we know on other (widely accepted, incontrovertible) grounds that w3's being better than w1 is what shows that a1 is wrong.

So: as to outcome comparison, not recognizing w3 means that we can't rank w2 as better than w1. And, as to act evaluation, not recognizing w3 means that a1 and a2 are both permissible. Hence our arms are tied—we can't solve the nonidentity problem—if we accept the axiological constraint.

[38] The person-affecting intuition and the rejection of transitivity of the *at least as good as* relation between outcomes are often considered to come together. In fact, however, they come together *only if* the person-affecting intuition is formulated by reference to the axiological constraint (which might, e.g., instruct in the context of the mere addition paradox that A+ is at least as good as A) rather than by reference to modally sensitive principles (which will instead leave room for the result that A+ < A).

I would resist that position and indeed that understanding of the preceding discussion. It's a rush to judgment. We should instead first examine whether the telic component of the person-affecting intuition really is subject to the axiological constraint—whether, that is, the classic view really does squeeze the person-affecting intuition into the mold of PAIO(c) and rule the modally sensitive PAIA* out-of-bounds from the start.

(ii) There is a second, and final, option for preserving the deontic person-affecting solution to the nonidentity problem while accepting PAIO(c) in place of PAIO*. We could reject the very tight connection between the evaluation of acts and the evaluation of outcomes. We could then insist that a1 at w1 is wrong even though w1 ≥ w2 ≥ w3.

But that option seems unpalatable as well. For one thing, it seems clear, on grounds that have nothing to do with the nonidentity problem, that PAIO(c) is objectionable. Surely it's not the case that w1 is at least as good as w3 is.

Moreover, if we reject the very tight connection between the evaluation of acts and the evaluation of outcomes, then the questions why we are interested in the latter at all and what we are supposed to look for in the successful candidate immediately arise. I'm not sure, however, that we can answer those questions.

3.5.4   *How the axiological constraint undermines narrow neutrality.* PAIO(c) doesn't just mean that we can't effectively defend the person-affecting intuition against the nonidentity problem. It also means that we must abandon the closely related happy child half of the *procreative asymmetry*—the idea, that is, that

---

Moreover, there is good reason to think that the well-formulated person-affecting intuition actually *rescues* transitivity. Thus the person-affecting intuition—related, as it is, to the intuition of narrow neutrality; see part ___ below—rejects the notion that the existence of the additional happy person makes an outcome better. Once we reject that notion, we are in a better position to avoid the repugnant conclusion and still insist on transitivity.

For discussion of the relation between transitivity and alternative forms of consequentialism, see Carter 2015 [in AAA Research].

it doesn't make things better to bring an additional happy person into existence—and specifically what I will call *narrow neutrality* in Part II.

Suppose that we want to compare w1 against w2. PAIO(c), as we have just seen, would instruct that w1 ≥ w2.

Nonidentity problem aside, on its face that may not *seem* an alarming result. But in fact it is a *quite* alarming result—*if* we want to retain narrow neutrality, according to which w2 (in which p never exists) ≥ w3 (in which p exists and p's wellbeing is maximized). On that principle, the addition of p in w3 doesn't make w3 better than w2. (Here the principle is plausibly *narrow*; unlike the neutrality intuition itself, that is, what Broome calls the neutral *range* claim, *narrow* neutrality leaves open that the addition of p *in w1 does* make w1 worse than w2). Transitivity, again, tells us that w1 ≥ w3, a result that, as noted just above, seems clearly false. Our only palatable option—if, that is, we keep PAIO(c)—may then seem to be to reject narrow neutrality—to accept, that is, that the existence of the additional happy person *does* make a given outcome better. But we don't want to do that—hence the bind.

3.6 *Evaluating the axiological constraint*

We thus need to get out of the straightjacket that the axiological constraint imposes—not just for the sake of the person-affecting intuition but for the sake of narrow neutrality as well.

But is that really so very hard to do?

3.6.1 *My proposal.* Suppose our interest is to compare w1 against w2 in a case where w3 exists as an accessible outcome. I want to be able to say that relevant to the evaluation of how w1 compares against *w2* is the fact that w1 makes things worse for p than *w3* does. I want to say that w3's accessibility tells us that w1 is "bad for" p, in the morally relevant sense that it is *worse for* p than is at least some other accessible outcome.

PAIO*, of course, lets us take that fact into account and conclude that the principle's necessary condition on one outcome's being worse than another is

satisfied. The necessary condition on w1 being worse than w2 is satisfied, not in virtue of the fact that w1 is worse for an existing or future person p than *w2* is—it isn't—but rather in virtue of the fact that w1 is worse for p than *w3* is.[39]

But as we've seen what we would *like* to be able to say runs afoul of the axiological constraint, which itself, as we have also seen, arises out of the effort to squeeze the telic component of the person-affecting intuition into the classic view. The axiological constraint insists that the person-affecting necessary condition on when one outcome is worse than another must be spelled out in a certain way— spelled out, specifically, in a way that approves PAIO(c) but immediately rules PAIO* out-of-bounds.

The question we now face is whether the axiological constraint really is a constraint we have no choice but to accept.

But how we answer that question, in turn, depends on how we answer two others: is the classic view itself a view we have no choice but to accept, and, if it is, does the classic view generate the axiological constraint?

3.6.2 *Inconsistency argument for the classic view*. Arguments for the classic view may seem plentiful. But some of those arguments aren't persuasive and others aren't fully developed.

Thus we noted earlier that on the classic view it is irrational to allow considerations relating to wγ—those *modal details*—to affect—that is, to *change*— how we rank wα against wβ. The basis for that judgment of irrationality was that when the goal is to determine whether an outcome is *better*—and not to determine whether an outcome is *the best*—a correct procedure for making that determination *need* not insist that we look beyond wα and wβ themselves. And if we need *not* look beyond wα and wβ, then it would be irrational to think that looking beyond wα and wβ could change our result. And that may be. But note that from none of

---

[39] We're not making the judgment that p is *wronged* in w1; that comes later; that's not what grounds the result that w1 is worse than w2 is. We're just noting a fact not about worseness between outcomes, but about one outcome's being worse *for a person* than another outcome is.

that does it *follow* that, for the purpose of comparing wα against wβ, facts about wγ can have *no bearing* on our evaluation. For it remains at least possible that facts about wγ—details that may bear on how wα is to be ranked against wβ—are already reflected, *sotto voce*, in wα and wβ. But if *that's* how the classic view is itself to be understood—as, that is, open to that possibility—then it doesn't generate the axiological constraint.

Other arguments may seem to support a stricter reading of the classic view. Such arguments may cite the axiom of the independence of irrelevant alternatives— for short, the *independence axiom*. According to that principle, how wα compares against wβ is not itself affected by the existence of wγ as a further accessible alternative; if wα is worse than wβ when wγ *is* an accessible outcome in the particular case, then wα is worse than wβ when wγ *isn't* an accessible outcome. It may well seem natural enough—though we shall come back to this point in what follows—to read the principle as ruling out the possibility that facts about wγ— *even if* already reflected, *sotto voce*, in wα and wβ—cannot make a difference to how wα compares against wβ. However, on that reading of the principle, there's very little light between the principle on the one hand and the axiological constraint on the other. So to cite the one in favor of the other isn't persuasive.

Still other arguments might simply appeal to a certain concept of intrinsic value or just to the position—adopted, i.e., by Temkin as critical to the *intrinsic aspect view*—that, it being wα and wβ we aim to compare, we should have no need to look beyond wα and wβ to make that comparison.

I think, however, that the best argument in favor of the classic view, *strictly* construed—construed, that is, so that it *does* support the axiological constraint—is an *inconsistency argument*. Specifically, it's the argument that, without the classic view, strictly construed, we will end up with an inconsistent ranking of outcomes. We'll end up saying that two outcomes *aren't* equally good in a case where the third outcome *is* accessible and that the same two outcomes *are* equally good in a case where that third outcome *isn't* accessible.

We've already introduced a schematic for the three outcome case (Graph 3.5.1). To see how the inconsistency argument works, we need a schematic for the two outcome case as well.

Graph 3.6.2: Two Outcome Case		
wellbeing	w1 (including a1)	w2 (including a2)
+10		
+5	**p**	
+0		*p**

Now, for the three outcome case, we want to say that w2, where p never exists, is equally as good as w3, where p's wellbeing is maximized, and that w1, where p's wellbeing is avoidably *not* maximized, is worse than both. And, as we have seen, we can say just those things under PAIO*.

But for the two outcome case, we want to say that w1 is equally as good as w2 is. And again we can say just that under PAIO*.[40]

But now we have an inconsistency. The modally sensitive PAIO* seems to instruct that both that w1 is equally as good as w2 is and that w1 is worse than w2 is.

The classic view, strictly construed, guards against just such an inconsistency. If w1 is worse than w2 in the three outcome case, then according to the classic view, strictly construed, w1 must be worse than w2 in the two outcome case as well.

---

[40] The necessary condition is failed, w1 not being worse for p than w2 is, and we infer that it's not the case that w1 is worse than w2 is. And we infer as well that it's not the case that w2 is worse than w1, w2 not being worse for anyone who does or will exist in w2 than w1 is.

I think, however, that this argument to inconsistency fails. Consider, again, any case that fits the three outcome schematic. In any such case, it's clear there will be a causal explanation of w3's accessibility—an explanation rooted in the *modal details* inherent in w1; an explanation rooted in how things, consistent with the laws of nature[41], *could* have been. Specifically: agents in w1 had the power, the ability, to bring w3 about and thus make things better for p and they declined to exercise that power. Now consider the case where w3 *isn't* accessible—the two outcome case. There, the reason w3 isn't accessible can also be explained by reference to what is going on in w1. To say that w3 isn't accessible is to say that agents in w1 *lacked* some power, some ability, to make things any better for p than things are in w2.

But that agents have a certain ability in the one world and lack that ability in the other just means that those worlds—w1 in the three outcome case and w1 in the two outcome case—are actually two *distinct* worlds. Worlds, after all, aren't simply *distributions*—bare boned assignments of wellbeing levels to members of a particular population. Rather, worlds come to us with all their details necessarily intact. New details entail new worlds.

And that in turn means that the inconsistency itself is illusory. It's one we immediately avoid upon the introduction of a more exacting vocabulary.

Thus we might say, about the three outcome case, that w1 is worse than w2 is and that w2 is exactly as good as w3 is, and, about the two outcome case, that w1′ is exactly as good as w2′. For the sake of completeness, we can even add that w1′ and w2′ are equally as good as w2 and w3 are, and that w1′ is better than w1 is (that last point, *despite* the fact that the two worlds distribute wellbeing across identical populations in identical ways).[42] These are all perfectly consistent results. No inconsistency or failure of transitivity emerges from anything we have just said.

---

[41] And (perhaps) the acts of other agents.

[42] The position that worlds may be distinguished on the basis of their accessible alternatives is not original. [cite] But to my knowledge the point hasn't been discussed as an element critical to any rescue of the person-affecting intuition from the threat posed by the classic view and, specifically, the axiological constraint.

3.6.3 *Implications for the classic view; independence.* It's worth noting that we haven't here relativized our comparisons to particular cases or choice sets. That means that we remain free to do just what we've done in the foregoing paragraph—complete the ranking, and compare not just the accessible outcomes in the one case against accessible outcomes *in that one case* but also compare accessible outcomes in one case against accessible outcomes in *another* case.

If the classic view means that, in comparing w1 against w2 in the three outcome case, we need not and indeed *must* not look beyond w1 and w2 and need not and indeed *must* not look at *any facet of w1 or w2 that will tell us whether w3 is an accessible outcome or not*, then we should give up the classic view. But I see no reason why we need to understand the classic view so strictly. Rather, we can instead understand the view to include the idea that we can complete the comparison of w1 against w2 *provided* that we have scrutinized w1 and w2 *closely enough to determine* what the powers and abilities of the agents in w1 and w2 in fact are and, specifically, to determine whether or not p's existence in w1 represents the best agents could have done for p.

In other words, it's the features inherent in w1 and w2—the *modal details* that are part of the very *identities* of w1 and w2—that decide whether w3 exists as an accessible outcome or not. To say that w3 exists as an accessible outcome for the relevant agents is just shorthand—and a convenient and perspicacious shorthand at that—for our saying *something about w1 and w2*. Thus we can after all retain the classic view—*so understood*—but still reject PAIO(c) in favor of PAIO*.

A similar point holds for the independence axiom. We need not understand that principle to say that how we rank w1 against w2 is independent of any facts having to do with w3 *even if* those facts are themselves reflected in w1 and w2. We can instead understand the principle to say that how we rank w1 against w2 *may well* depend in part on facts having to do with w3 *insofar as those facts are reflected in w1 and w2*. Which is just to say that w3 is in very real sense independent to our comparison of w1 against w2, but independent *only* because we find whatever facts

about w3 that we need to rank w1 against w2 within the confines of w1 and w2. Looking *carefully* at w1 and w2 is, in other words, precisely the same operation as taking w3 into account.

The upshot? *All that really must go is the axiological constraint itself*, the product of the thought that the only way to squeeze the person-affecting intuition into the classic view is for the necessary condition on one outcome's being worse than another to be satisfied only if the one outcome is worse for p than the other outcome is *for p*.

3.6.4 *Accessibility axiom.* Residual concerns about inconsistency are addressed by the following principle.

> *Accessibility axiom.* If wβ is accessible to wα, then *necessarily* wβ is accessible to wα.

The accessibility axiom insures that the way of reconciling the results we spelled out for the three outcome case—w1′ is worse than w2′—and for the two outcome case—w1 is equally as good as w2—will hold for any relevantly similar pairs of cases. It guaranties that we won't stumble across still *another* pair of cases where w1′ in the two outcome case turns out to be identical to w1 in the three outcome case. If that third outcome w3 turns out to be *inaccessible* in a given case, in other words, it won't be *w1* that that particular case involves but rather some other outcome altogether.

That a non-standard—that is, person-affecting, or modal—approach relies on the accessibility axiom should not be viewed as itself problematic. After all, the accessibility axiom itself is highly plausible, really perhaps just a product of the concept of accessibility in combination with a very basic understanding of what a possible world consists in.

3.6.5 *Addition plus and the axiological constraint.* Supporting the argument that we should discard the axiological constraint in favor of a more far-reaching, modally sensitive way of comparing outcomes is *addition plus*.[43]

---

[43] This variation on Parfit's mere addition paradox, in addition to challenging the axiological constraint itself, makes many other points as well. Thus addition plus, like the miserable child half of the asymmetry, nicely shows that some formulations of the person-affecting intuition—e.g., *modal actualism*—will not work. (If w1 is actual, it won't do, e.g., to say that only the actual person p matters morally—that it matters not at all how well off q is in w2 and w3 and hence, e.g., that a1 is wrong or that w1 is worse than w2.)

wellbeing	w1 (including a1)	w2 (including a2)	w3 (including a3)
		Graph 3.6.5: Addition Plus	
+10		**p**	
+9	**p**		
+5			**p, q**
+1		**q**	
+0	*q**		

When we compare w1 against w2 without taking w3 into account—without, that is, examining w1 and w2 closely enough to see that reflected in w1 and w2 is the fact that w3 exists as an accessible outcome—my view is that we really have not given ourselves enough information to make the comparison. And if we *do* proceed to make the comparison *without* getting that information—without, that is, coming to understand that w3 exists as an accessible outcome—we may well come to a result that anyone who wants to retain the person-affecting intuition can reasonably reject: that w2 is at least as good as w1 is.[44] To accept that w2 is actually—given w3—worse than w1 is to reject the axiological constraint. To accept the axiological constraint, on the other hand, may well be to give up the person-affecting intuition.

In contrast, consider how the classic view—strictly construed—enables Lazari-Radek and Singer to move without discussion from the result, in a case they have presented as a two outcome case ("natural phenomena . . . cannot be changed and . . .compensation [for q] is impossible") that w2 is at least as good as w1, to the result, in a case that they then expand to include a third outcome w3, that w2

_____

[44] Thus PAIO* allows us to say about this case that w2 is worse than w1 in view of the fact that w2 is worse for q than *w3* is. At the same time PAIO* insists that—in the two outcome case—w2 is at least as good as w1 is. This inconsistency—as noted earlier—we can resolve by distinguishing w1 from w1′ and w2 from w2′.

remains at least as good as w1.[45] But the inference that the classic move—strictly construed—on its face seems to support is questionable. In the two outcome case, the natural phenomena "cannot be changed"; compensation is "impossible"—that's why it is a *two* outcome case. In their three outcome extension, it must be that w2 has now been stripped of those facts; now, phenomena *can* be changed, compensation *is* possible, else it wouldn't be a *three* outcome case.

If, however, we take the scenarios under consideration to be worlds, and understand worlds to have their properties necessarily, then it seems that one case or the other—the two outcome case, or the three outcome case—must be dismissed as impossible.

But there's a more substantive point to be made as well. It seems perfectly consistent and reasonable to say that what we want to say about the first two outcomes is just going to vary, depending on whether the third outcome exists as an accessible outcome or not. And if that point itself is consistent and reasonable, then the door is open to our rejection of the axiological constraint. Having rejected that constraint, we are then in a position to put PAIO* forward as a properly formulated version of the person-affecting intuition, in place of PAIO(c).

3.7 *Remaining work.* Of course, cautions that we should not assume that the person-affecting intuition says one thing—PAIO(c)—when in fact it's better understood to say another—PAIO*—piled on top of cautions that we should be alert to the modal details of our own nonidentity cases doesn't *insure* a solution to the nonidentity problem.

If the nonidentity case doesn't come with any critical modal details—if, that is, our modally sensitive formulations of ontic and telic versions of the person-affecting intuition don't have anything much to see when they do their requisite looking around to take into account all the facts of the relevant case—then the nonidentity problem will remain unsolved. Indeed the inconsistency that the problem ends with—that a1 both is and isn't permissible—would mean that we must put the person-affecting intuition itself back on the chopping block.

---

[45] Lazari-Radek and Singer 2014, pp. 366 and 371.

The task now, then, is to show that the most challenging form of the nonidentity problem comes with the relevant modal details. The task now, in other words, is to show that, when the act under scrutiny is clearly wrong, we can explain by reference to those modal details just why the person-affecting intuition avoids the result that the act is permissible.

Chapter 4
Kavka's Pleasure Pill Case

4.1 *Avoiding the mistakes*

Kavka's pleasure pill case is the perfect exemplar of the nonidentity problem in its most challenging form. We are—and, contra Boonin, remain—confident that the act under scrutiny is clearly wrong. At the same time, it's not clear at all to us starting out that that act performed at the particular world has made the child whose plight is so concerning to us to begin with or anyone else who does or will exist in that world any worse off. Rather, it seems to us starting out as though the child surely *owes his or her very existence*—itself an existence worth having—to the act under scrutiny. That means that the necessary condition the person-affecting intuition sets forth, in both its ontic and its telic forms, is immediately failed. The choice is permissible; the one world isn't worse than the other.

To untangle the argument, we must avoid Possible Mistakes A and B. We must present the case in a way that explicitly recognizes its *critical modal details*. And we must apply the *modally sensitive* formulations of both the ontic and the telic components of the person-affecting intuition and avoid the *unduly constricted* formulations. We can then defend the only version of the intuition that we really have any interest in defending to begin with, the version that puts those modal details to work to work in a way that avoids the inconsistency.

4.2 *Modally enriched presentation of the facts*

Thus suppose that an agent, Luc, prior to conceiving a child, pauses to take a *pleasure pill*—a pill that is teratogenic but that produces a mild and transient euphoria in Luc. Luc then proceeds to conceive a child, Andy. Andy's life is clearly worth living. But his life is burdened as a result of the impairment and his overall lifetime wellbeing accordingly reduced.

Now, as Kavka himself notes, had the agent *not* paused to take the pleasure pill—had he, e.g., taken an aspirin instead—the very same child *might* still have

been conceived.[46] But—given the very low probability that any one person would ever have come into existence at all had things been other than just the way they in fact were; given, that is, the *precariousness of existence*—we recognize that the odds are very much against Andy's coming into existence had Luc taken the aspirin rather than the pleasure pill.

Indeed, we make it part of the case that, had Luc not paused to take the pleasure pill, the timing and manner of conception would have been different and Luc would have conceived a distinct child in place of Andy. (Perhaps he rushes to take the pleasure pill and would have taken his time getting to the aspirin.) Let's stipulate further that that distinct child—say, Ruth—would have been better off than Andy in fact is—that is, that Ruth is better off at the closest possible world where Luc *doesn't* take the pill than Andy is at the actual world where Luc *does* take the pill.

Now, this isn't, as we shall see, a *complete* presentation of the pleasure pill case. There's an important probability point we've yet to make. But let's sum up what we have so far:

---

[46] Kavka 1982, p. 100, n. 15.

Modally Enriched Nonidentity Problem (Incomplete)

Let w1 be the actual world. Let a1 performed at w1 be Luc's act of taking the pleasure pill at w1. Let Andy be a child born seriously impaired at w1 as a result of Luc's taking the pleasure pill at w1.

Let w2 be an alternate accessible world. Luc performs a2 at w2 in place of a1 at w1. Let Ruth be a child nonidentical to Andy born healthy at w2 as a result of Luc's performance of a2 at w2.

We stipulate that w2 is better for Ruth than w1 is for Andy—that is, that Andy has less wellbeing in w1 than Ruth has in w2. We also stipulate that, while Andy's wellbeing is reduced as a result of the impairment, his life at w1 is clearly worth living—that is, that his wellbeing is clearly in the positive range.

Let w3 be a further accessible world in which Luc performs a3 rather than a1— that is, pauses to take the aspirin rather than the pleasure pill—and nonetheless conceives Andy. w3 is better for Andy than w1 is.

Let's recognize that, while w3 is *technically* accessible just prior to performance, the *probability* of Luc's conceiving Andy, given that Luc chooses to take the aspirin rather than the pleasure pill—the probability, that is, of w3 obtaining, given a3—is very low.

Since Andy never exists in w2 and has a life clearly worth living in w1, we infer that it's not the case that w1 is worse for Andy than w2 is. We also stipulate that no one other than Andy who does or will exist in w1 is affected by what the agent does. Hence we infer that it's not the case that w1 is worse for anyone who does or will exist in w1 than w2 is.

We stipulate, finally, the following counterfactual: had Luc not performed a1 at w1, Luc would have performed a2 at w2 and Ruth would have been the one conceived rather than Andy.

[END]

Or, in graph form:

wellbeing	w1 (including a1)	w2 (including a2)	w3 (including a3)
+10		**Ruth**	**Andy**
+8	**Andy**		
+0	*Ruth**	*Andy**	*Ruth**

Graph 4.2: Modally Enriched Nonidentity Problem

Thus the facts of the pleasure pill case presented in a modally *enriched* framework. Though incomplete, it's a plus that this modal presentation of the facts explicitly recognizes that the future w3 in which Andy both exists and is better off than he is in w1 is accessible to agents. We thus avoid the mistake of ignoring some of the facts of our own case.

4.3 *Objection to modal presentation*

The critic of the person-affecting intuition might object that the details we've added to the case as we work toward completing the presentation unfairly *weaken* the nonidentity problem. The critic might object that we could just as well have added details that instead *strengthen* the problem.

Specifically, rather than completing the presentation by stipulating w3 as an accessible outcome, the critic might object that we could have completed the presentation by *stipulating* that *no such accessible w3 exists*—by *stipulating*, that is, that the outcomes accessible to Luc in the pleasure pill are *exhausted* by just w1 and w2.

Two notes here. (i) One *might* attempt to set w3 aside as *irrelevant* to the evaluation. One *might* have thought such an attempt sensible in virtue of our concession that w3 is highly improbable or our stipulated counterfactual to the effect that, had Luc no performed a1 at w1, he would have performed a2 at w2. We

explored and questioned those rationales earlier. But still it's understandable that one *might* have considered the strategy valid.

In contrast, it seems *undeniable* that w3 exists as an *accessible* outcome—however merely technical; however improbable; however counterfactually doomed—for Luc in the pleasure pill case. A world is *accessible* if it's a possible future the agent has the *ability* to bring about; a world is *accessible provided* it's *not* barred by the laws of nature or by the acts of other agents.

Suppose I need to open a safe to get to a bomb that I can then disarm and which will otherwise blow up the building. But I don't know the combination to the safe. Twirling the dial this way and then that and succeeding in opening the save is nonetheless part of a possible future that is *accessible* to me. I *can* open the safe, even if, very probably, the combination I randomly try *won't* be the right combination.[47]

Ditto the pleasure pill case. There is nothing in Luc's switching from the pleasure pill to the aspirin that would renders it contrary to the laws of nature or acts of other agents for Luc then to conceive Andy. The pleasure pill *isn't* a fertility drug.

This is not to say, of course, that Luc's performing a3 would *guaranty* that w3 would obtain. Rather, including w3 as an accessible outcome in the presentation of the facts of the pleasure pill case simply makes explicit our background understanding that, among all the ways there are out there of Luc's implementing the choice to take the aspirin rather than the pleasure pill, there exists at least one such way—one such act; call it *a3*—that will form part of a chain of acts and events that will yield an outcome—which we call *w3*—in which Andy is conceived and is better off than he is at w1. There is at least one such act, that is, *a3*, that *mimics* a1 in all those spatial-temporal-mechanical features that play any causal role at all in Andy's being conceived rather than, e.g., Ruth, and there is at least one such world in which things then unfold at that world *just as they unfold* in w1.

---

[47] [Cite for this example.]

The *timing and manner* of Luc's conception at w1 isn't, in other words, *unique* to Luc's performing a1 at w1. The *timing and manner* of Luc's conception *might* be *exactly* as it is in w1 even if a3 replaces a1.

Now, this way of thinking about the pleasure pill case presumes—as noted above—that *the laws of nature are as we understand them to be*. It presumes, in other words, that we come to the table with certain background facts in hand and that those background facts are fair game as we process the case and make the judgments that we then make about the case—including the judgment that a1 is wrong. We understand that not *every* historical blip is actually an essential, or necessary, ingredient to a given person's ever being conceived at all. Plausibly, you couldn't have had genetic parents other than the genetic parents you in fact had.[48] But you could have been conceived in Massachusetts rather than Oklahoma; and Andy could have been conceived had Luc taken the aspirin rather than the pleasure pill. Being conceived in Oklahoma thus was an *accessible* outcome for you. And being conceived Luc's having taken the aspirin rather than the pleasure pill was an *accessible* outcome for Andy. And that's all so, however highly improbable it is that that better-for-Andy accessible outcome would have obtained had Luc not done just as he did.

Can the objection be pressed harder? It's, after all, the critic's hypothetical to do with as he or she pleases. Is it legitimate to present the pleasure pill as just as we have—and then to add the stipulation that after all w3 *isn't* an accessible outcome?

Actually, no—at least, not without a lot of further work. To stipulate that the pleasure pill case is a *two outcome* rather than a *three outcome* case is at odds with our coming to the table with the understanding that the laws of nature are as we understand them to be. It introduces an ambiguity into the case. That ambiguity

_____

[48] One clarification. "Couldn't" here doesn't mean *logically* couldn't, but rather that the technology to make it happen isn't *available* to agents at this point in this world. However, your having genetic parents other than the genetic parents you in fact have would, in addition, seem to be a logical impossibility. I am here assuming—and take as plausible—the genetic origin theory of personal identity.

renders the judgments we make about the case unreliable—whatever our subjective level of confidence might be. And the case thus fails as a counterexample.[49]

The critic can always to a more thorough revision of the case—one that insures that the stipulation that eliminates w3 as an accessible option is *consistent* with the background that we come to the table with, including the presumption that the laws of nature are as we understand them to be; one that insures that we are not trying to have things both ways.

Thus the critic might stipulate, not just that it's a two outcome case, but also that the laws of nature *aren't* as we understand them to be. Thus the critic might have us imagine that Luc takes the teratogenic pleasure pill in a world w1-*alt.* where the pleasure pill acts just like a fertility pill does at the *actual* world. In that far-away world w1-alt., the pleasure pill indeed imposes certain risks on any offspring

---

[49]   Cases that effectively counterexample the person-affecting intuition or any other principle cannot smuggle into their facts ambiguities regarding those facts that, for all we know, *destabilize* the judgments that we make about those facts—the very judgments that are meant to show that the particular principle that is being tested is false. But the facts that are included in a given case not limited to what is explicitly said. They include a background of further facts that is itself rooted in our understanding of how the world works—our presumption that the laws of nature are as we understand them to be. Now, we are perfectly free to stipulate that the laws of nature are other than they in fact are. We can make up whatever hypothetical we want, provided just that the hypothetical is itself logically possible and conceptually coherent. What we can't do is present things as though the laws of nature are just as we take them to be—and then stipulate that the laws of nature are after all nothing like we take them to be. I am grateful to Adam Lerner for discussing this general point with me in connection with the trolley cases; [add cite].

Thus, for the sorts of nonidentity cases that I focus on here—the most challenging cases; the cases in which we are confident that the act is *clearly wrong*—we can always stipulate that w3 isn't part of the case. But we do that, we must also make it explicit that it's also part of the case that the laws of nature are other than they in fact are. We can always stipulate that w3 isn't part of the case—but if we do it, we must do it *consistently* throughout the case.

The problem is that as soon as we do that we badly weaken the case. The two outcome form of the nonidentity problem just doesn't challenge the person-affecting intuition as effectively as the three outcome form of the problem does. We can't be as confident that the act under scrutiny is itself clearly wrong. Transforming the pleasure pill into a fertility isn't a minor change in the case. As noted in the text that follows, it, rather, converts the case from one in which the act is clearly wrong to one in which it's not clear at all that the act is wrong.

then conceived. But it's also a pill Luc *must* take if he is ever to conceive any child at all. In that new case, and relative to that new world w1-alt., the world where Luc takes the aspirin instead of the pleasure pill and Andy is conceived *is* inaccessible. But in that new case it's surely at least *unclear* to us that what Luc has done is wrong. Before condemning his choice, we would surely at the very least want to know more about the risks involved.[50]

Thus the act under scrutiny in the new case doesn't meet the *clearly wrong* standard that—I would be the first to concede—the act under scrutiny in the pleasure pill case clearly meets. (Ditto the slave child case, the risky policy case, the depletion case and so on.)

(ii) What makes the nonidentity problem important isn't undermined by our recognizing w3 as an accessible outcome. Nor does that recognition take us very far at all in solving the problem. It's a step, but it's a baby step.

What makes the nonidentity problem such an important problem, rather, happens *after* we acknowledge w3 as an accessible outcome. As the medical examples we discussed earlier suggest, most of us consider the *probabilities* of a given case important to moral evaluation. We consider the calculation of *expected* value of the alternative acts critical to moral evaluation. What makes the nonidentity problem hard, then, is to say how an act can be wrong in virtue of its being bad for a particular person *notwithstanding* the fact that *any alternate act* reduces the probability that *that* person will ever come into existence at all.

We turn to that point now.

4.4 *Probability*

---

[50] This is not to say that *any and all* fertility treatments are permissible. Those involving supernumerary pregnancies are, e.g., notably problematic. But, interestingly, in those cases, for each of the surviving infants, the burden imposed on that infant was not *unavoidable* at all. For each such infant, in other words, there exists an accessible outcome in which that infant was better off. For each such infant, the agents had the ability to avoid the burden on behalf of *that* infant by reducing the pregnancy earlier on by way of selectively aborting some of the *other* developing fetuses.

As noted, the above presentation of the pleasure pill case, though modally enriched, isn't itself complete. We are still missing some facts, and, in particular, some facts relating to *probability*.

Now, the modal presentation explicitly includes the concession that the probability of Luc's conceiving Andy, given that Luc takes the aspirin rather than the pleasure pill, is very low. Let's just note that there are actually two hurdles that must be overcome in order for the probability of Andy's coming into existence to be anything more than very low, in the case where Luc chooses to take the aspirin. First, having chosen to take the aspirin, Luc must then *implement* that choice *by a3* and not by any of the *many* alternative acts that would equally well implement his choice to take the aspirin. He must, that is, implement his choice by an act that, like a3, mimics a1 in its various spatial-temporal-mechanical respects. And, second, from the point at which the implementing act is itself completed until the point at which conception takes place, the future must unfold in just the way that it does in w1. If, that is, upon the completion of a3, Luc then forbears ejaculation until he returns from a trip around the world—if, e.g., such a *w4* unfolds in place of w3—and, upon the completion of a1, Luc proceeds immediately to intercourse, Andy won't exist.

That's a lot of uncertainty. That uncertainty is built into our modal presentation. But what *isn't* included there is just as important: *that those same hurdles are in place in the case where Luc chooses to take the pleasure pill*. There, too, having chosen to take the pleasure pill, Luc must then implement that choice *by a1* and not by any of the *many* alternative acts that would equally well implement his unfortunate choice. And there, too, having completed the implementing act, the future must unfold in just the way that it does in w1. Luc can't that is, take the pleasure pill and then—departing from the future that in fact unfolds in w1—take the trip around the world and still conceive Andy.

Now, some theorists do not think that the probabilities matter in the context of moral evaluation. Such *actual value* consequentialists will focus strictly on what Luc could have done (whether, specifically, he could have done more for Andy than he in fact does at w1). Pertinent to their analysis will be *accessibility*.

But for those theorists who—like Kavka—think that the precariousness phenomenon bears on what the person-affecting intuition tells us about the case, *probability* is relevant. What I have argued here, however, is that the precariousness phenomenon cuts both ways. The probabilities involved in the case—whether Luc takes the pleasure pill or the aspirin—are a wash. Expected value theorists, in such a case, will look to the actual value of each of the (equally low-probability) outcomes to determine what the agent ought to do. And in this case that actual value that w1 assigns to Andy is lower than the actual value w3 assigns to Andy.

To complete, then, the presentation, we thus need to include the following addendum:

Addendum to Modally Enriched Nonidentity Problem

Similarly, while w1 is technically *accessible* just prior the choice, the *probability* of Luc's conceiving Andy, given that Luc chooses to take the pleasure pill, is very low (and is indeed no greater than the probability of Luc's conceiving Andy, given that Luc chooses to take the aspirin.).

4.5 *Modally sensitive principles*

We've moved beyond the modally *impoverished* framework reflected in standard presentations of the nonidentity problem and now situated the case in a modally *enriched* framework.

We've avoided, in other words, Possible Mistake A. To avoid Possible Mistake B is just to make sure that we now apply the *modally sensitive* formulations of the person-affecting intuition in place of the *unduly constricted* formulations of the intuition. Thus we abandon the clearly false counterfactual PAIA(c).

The latter is an especially critical step. It's our rejection of PAIA(c) that allows us to identify the counterfactual stipulation in the modal presentation of our case as the red herring that it is. It's a stipulation of the case—we don't challenge it; and we recognize that, unlike the two outcome stipulation, the counterfactual stipulation is certainly *not* at odds with our understanding of how the world works.

And at first glance it might seem highly relevant to the moral evaluation of a1. But we are now in a position to point out that that stipulation is relevant to the evaluation *only if* we retain the false PAIA(c)—which we, of course, don't want to do.

The revised argument thus will rely not on the false PAIA(c) but rather on the far more plausible PAIA* and PAIA**.

Making that change requires, of course, that we make further conforming changes throughout the argument else we lose validity. The claim must now be, not that a1 at w1 is better for Andy than things otherwise *would* have been, but rather that a1 at w1 is better for Andy than things were under each other alternative act performed at each other accessible world.

Summing up:

Line no.	Revised Nonidentity Argument/Deontic Form	Justification
1	a1 performed at w1 is morally wrong.	Intuition
2′	There exists no alternative act aβ performed at any alternative accessible outcome wβ such that a1 at w1 makes things worse for Andy (or anyone else who does or will exist at w1) than aβ performed at wβ does. [Alternatively: There exists no alternative act aβ performed at any alternative accessible outcome wβ such that the expected wellbeing of a1 at w1 for Andy is less than the expected value of aβ at wβ is for Andy (or anyone else who does or will exist at w1.]	Stipulations
3′	aα performed at wα is morally wrong *only if* there is at least some person p such that p does or will exist in wα and there is an alternative act aβ performed at an alternative accessible outcome wβ such that aα performed at wα makes things *worse for* that person than aβ performed at wβ does. [aα performed at wα is morally wrong *only if* there is at least some person p such that p does or will exist in wα and there is an alternative act aβ performed at an alternative accessible world wβ such that the expected wellbeing of aα performed at wα for p is less than the expected wellbeing of aβ performed at wβ for p.]	PAIA* [PAIA**]
4	It's not the case that a1 performed at w1 is morally wrong.	Lines 2 and 3
5	a1 at w1 both is and isn't wrong	Lines 1 and 4

But this revised argument can be quickly evaluated. Premise 2′, in both its forms, fails. There exists an available act and an accessible world such that that act *both* makes things better for Andy—that is, generates more *actual* wellbeing for Andy— *and* can be expected to make things better for Andy—that is, generates more *expected* wellbeing for Andy—than a1 at w1 does.

Hence the necessary condition set forth in both PAIA* and PAIA** is satisfied, and we never get to the result that a1 at w1 is permissible and thus never face inconsistency.

4.6 *The revised telic argument*

The argument involving the telic component of the nonidentity problem parallels the deontic argument and we won't bother charting it here. What is important to note is that we similarly avoid inconsistency since we never reach the result that it's not the case that w2 is worse than w1—that is, that w1 is at least as good as w2 is. While w2 isn't worse for Andy than *w1* is, w2 is worse for Andy than *w3* is. Accordingly, the necessary condition on outcome worseness that PAIO* sets forth is satisfied. We thus avoid inconsistency—and at the same time leave the door open for other person-affecting principles to instruct that w1 is, after all, worse than w2 is. The upshot? w2 and w3 are equally good—and w1 is worse than both.

Chapter 5
The Two Outcome Form of the Nonidentity Problem

The sort of nonidentity case I have considered here—that is, the three outcome problem—does not itself exhaust the nonidentity problem. We must also take seriously still another sort of nonidentity case—the two outcome problem, where it's simply part of the case that the agent really *can't* do anything more for the burdened child than the agent already has.

An example would be a case in which the agent conceives a child who will be burdened by a genetic condition that we today do not have the ability to cure or substantially mitigate. The child's life nonetheless will be clearly worth living. Moreover—and here it becomes evident that the world we are imagining is quite distant from our own actual world—that child's coming into existence will not make things worse than they are, for any person who does or ever will exist in the one world, in any other accessible world. Bringing the genetically burdened child into existence is, in other words, actually *and* expectationally *maximizing*—for each person who *does or ever will exist* at that world, including the child.

The principles I've ended with—PAIA* and PAIA**—both will deem the act under scrutiny permissible.

But I believe that that act is permissible. At least, it's not even close to meeting the *clearly wrong* standard that the most serious challenges to the person-affecting intuition (contra Boonin) clearly meet.

We *all* come into existence genetically burdened in some way or another; moreover, there's a lot of wrongdoing going around that's connected with procreation even under the best of circumstances. But it's not at all clear to me that, when we really are in a case where, for each and every person, whether existing or future, that person's actual wellbeing and expected wellbeing have been maximized, anything clearly wrong has been done.

Chapter 6

Conclusions

One of my primary aims here has been to bring to the surface certain details—*modal* details—of those nonidentity cases that give rise to what I believe are the most serious objections against the person-affecting intuition—cases, that is, in which the act under scrutiny is *clearly wrong*. I have made my points in connection with Kavka's pleasure pill case. But we can make exactly the same points for Kavka's slave child case, Parfit's depletion and risky policy cases and cases involving historical injustices and environmental failures, including climate change. Those modal details are often left unrecognized or—if noted at all—dismissed as irrelevant to the moral evaluation of the act under scrutiny. But either way a mistake has been made. Those details *are* part of the relevant cases—and *are* relevant to the moral evaluation.

Of course, what we accept as *relevant* to the evaluation is, in the end, a matter of what *principles* we have accepted as properly formulating the person-affecting intuition. But things pedagogically can happen the other way around. Thus the very details of the case—once recognized—can steer us away from problem principles—*unduly constricted* principles—and toward principles that do a better job articulating our underlying intuitions—*modally sensitive* principles.

Thus, whether we are thinking about the seemingly unique ethical issues that arise in *additional* person cases or perfectly ordinary *same* people cases (medical cases; the shoot-Harry-in-the-arm case), focusing on the alternatives beyond simply *what is* and *what would otherwise have been* helps us appreciate that standard formulations of the person-affecting intuition in both its deontic and its telic form will not do. We can understand that a *better* way of capturing what the intuition actually comes to is a less *constricted* way, a more modally sensitive way—a way that explicitly takes into account not just *what is* and *what otherwise would have been*, but *what could have been* as well.

Philosophers may have considered that seemingly obvious way of thinking about the person-affecting intuition to be ruled out-of-bounds in virtue of a certain axiological constraint. I have argued, however, that that constraint surely does not

have the reach they have assigned to it. It does not rule out an interpretation of the person-affecting intuition that requires us to look around and take into account *all* the facts of our own cases prior to evaluating a particular act as wrong or outcome as worse.

[END][Appendices omitted (see draft 2016.08.14 for appendices)]