

Counting the Costs Before They're Hatched:  
Climate Change, the Nonidentity Problem and the Moral Significance of Merely Possible Persons  
M. A. Roberts  
for  
Conference on the Ethics of Climate Change:  
Intergenerational Justice and the Global Challenge  
University of Delaware  
October 30-31, 2009  
robertsm@tcnj.edu

1. *The Nonidentity Problem: An Obstacle to an Accurate and Natural Assessment of the Damages of Climate Change*

The people who will endure the most severe consequences of climate change are, in the main, *future* people: people who do not yet, but will, exist. According to one important type of nonidentity problem, a large proportion of those same people will also “owe their very existence” to the choices that give rise to those consequences having been made in exactly the way that they were. Moreover, the lives of those people will be worth living. It isn't plausible that the effects of climate change will be so severe that never having existed at all will have been the better option; climate change does not mean ubiquitous *wrongful life*. These two points together suggest that many, even most, of the effects of climate change, going forward, will not be *losses* at all, but rather, if anything, *gains*.<sup>1</sup>

But we are, at the time, convinced that that picture of what climate change holds for future persons cannot be quite right. We are convinced, in other words, that the rosy picture the nonidentity problem paints vastly *undercounts* the costs of climate change. And undercounting the costs of climate change is a problem. For we need to have an accurate assessment of those costs, in order to figure out just how much we are willing to sacrifice, today, in order to avoid those costs on behalf of future persons. So we have a gap that we must fill in. We must figure out a way to explain why the effects of climate change are “bad” and why avoiding those effects is worth a great

deal of sacrifice on our part, despite the fact that, according to the nonidentity problem, many, or most, of those effects represent not losses at all but rather, if anything, gains.

I have little doubt that philosophers can fill in this gap if they must. But it is not a trivial project. We must set aside principles that would evaluate the choice to do nothing about climate change in any natural or intuitive way—on the basis of such straightforward inquiries as *how many people* can be expected to incur losses, now and later, as a consequence of our doing nothing today about climate change; *how deep* those losses will be; and whether those losses can somehow be *justified* on the ground that they are necessary if the imposition of still deeper losses on still other persons, including, e.g., existing persons, is to be avoided.

What principles do we put in their place? The most obvious approach would be to bring in *totalism*, the traditional utilitarian principle, the principle that evaluates choice by reference to total aggregate wellbeing. We then will not need to worry about whether climate change imposes any losses or not; we can live with the fact that *all* the victims of climate change “owe their very existence” to our choice to do nothing, since we will still obtain the result, under totalism, that the choice to do nothing about climate change is easily wrong even if it imposes no losses on any existing or future person at all. But it is worrisome that we happen also to be fully aware of that totalism faces grave challenges.<sup>2</sup> And we happen also to be fully aware that amending totalism in the direction of pluralism—that is, defining the “overall good” to include not just aggregate wellbeing but also such things as human flourishing and equality, all in an attempt to avoid the challenges that totalism itself faces—may well represent a promising strategy but not a strategy that has yet been implemented in any real detail, much less tested.

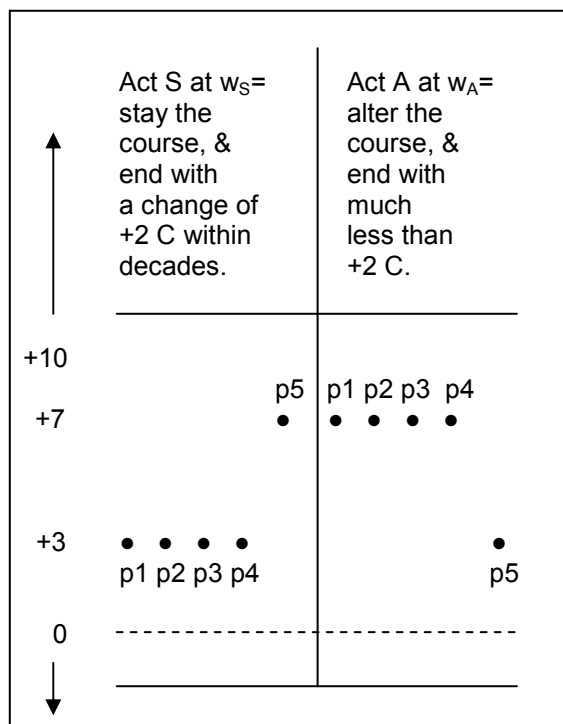
As moral philosophers, we feel that the world should be able to look to us for some insight on what, if anything, we ought now to do about climate change—what sacrifices we ought to make, what global and personal projects we ought to defer, why the failure to do something substantial to curb climate change is *not an option*. But what we have to say to the world seems a little weak.

Thanks to the nonidentity problem, and in the spirit of full disclosure, we can say that, from the perspective of future persons, *our failure is an option, and indeed it's their best option*. For any success we have in thwarting climate change means that the sweetness of even the flawed existence would never have been theirs at all. We can then fill in the gap, of course. We can argue that there exist grounds, beyond avoiding losses on behalf of future persons, for doing something substantial now about climate change. We can talk about totalism; we can talk about pluralism. We can, more generally, talk not about the option that makes things better for more persons, but rather the option that makes things better, in Sidgwick's phrase, "from the point of view of the universe." But can we really make up for the fact that the nonidentity problem trips us up rather badly, just as we are coming out of the starting gate?

In this paper, I want to propose an alternate way of getting ourselves out of this horrific jam. The nonidentity problem is, at this late date, an old problem, and I know that many philosophers feel that they just cannot endure still another discussion of it. All that can interestingly be said about the nonidentity problem has in fact surely been said. But I want to say one more thing about it here, briefly. And that is that the nonidentity problem—at least, the type of nonidentity problem we see at play in the context of climate change, and in Parfit's depletion and risky policy examples, and Kavka's slave child and pleasure pill examples, and in the context of discussions of historical injustices, including the Holocaust and the centuries-long practice of slavery in the United States—is not a problem but rather a mistake. It's a mistake, and a mistake we should at this juncture correct, so that we can clearly articulate, in an intuitive and natural way, the *full cost* of our choice to do nothing about climate change.

## 2. *The Argument to the No Harm Done Result*

Let's begin by setting up the problem.



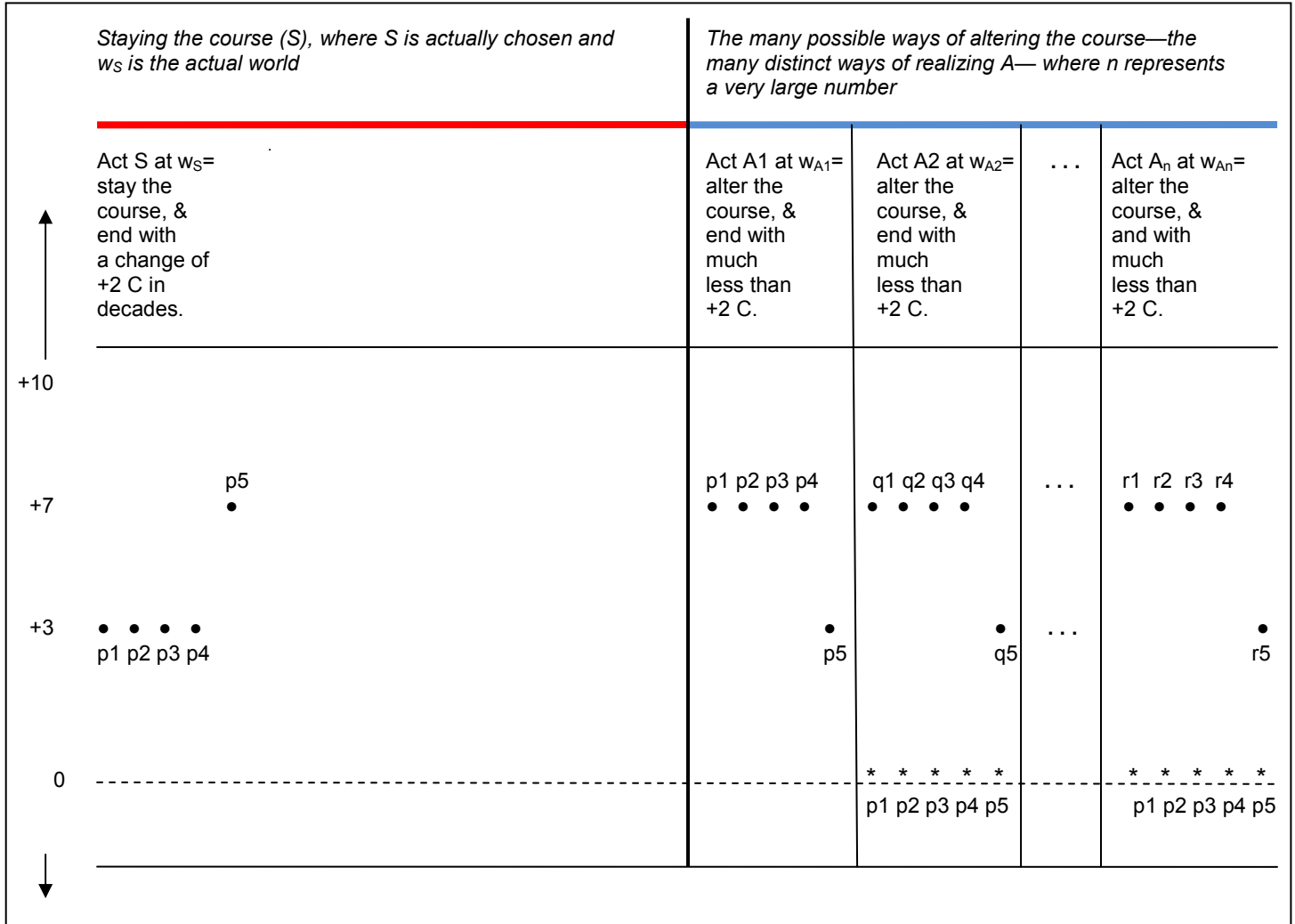
GRAPH I: NAÏVE ACCOUNT OF THE COSTS OF CLIMATE CHANGE

The nonidentity problem points out—and this part of the argument is perfectly correct—that this simple model for describing the effects of climate change on future persons is wildly inaccurate. Let's suppose that act S is in fact performed, and that world  $w_S$  is the actual world. The fact is that, for any person  $p$  who exists and suffers under S, the probability is *virtually nil* that  $p$  would have been any better off had we chosen anything other than S. The probability, instead, is overwhelmingly that  $p$  would never have existed at all. After all, had we adjusted our future-directed conduct *enough* so that the conditions of existence would have been predictably better for all or most of whatever people do later come into existence, including  $p$ , we would also have adjusted our future-directed conduct *so much* so that the timing and conditions of  $p$ 's conception would have been changed as well. But even the subtlest changes in the timing and conditions of conception for

Roberts, "Counting the Costs Before They're Hatched"

p are ample to insure that p would never have existed at all. Keep in mind here that any given male ejaculate contains hundreds of millions of such gametes. The chances that the very gametes that combined to produce p would have combined in the absence of S is practically nil. We are thus left with the nonidentity problem's most striking conclusion: the future-directed choice S that on the surface of things *looks* like it would be the most adverse to the interests of future persons is going to be the very choice that *in fact is* least adverse from the perspective of any future person. Our failure now —our performance of S—is, after all, *their* best option, going forward.

So, we fix the graph to reflect this perfectly correct point:



GRAPH II: CLIMATE CHANGE + NONIDENTITY

We leave the left hand side of the graph as it was before: we imagine that we have actually chosen to stay the course, and that the particular future persons who happen to come into existence under the particular way we happen to stay the course are just p1 – p5. But now we represent the *precariousness* of existence for p1 – p5 that inevitably would have come with our choice to have

done anything other than S. Many, many ways of realizing the "alter the course" choice—of realizing the choice of S—present themselves to us: A-sub-1 does that trick, but so does A-sub-2, A-sub-3, and so on, through A-sub-n, for some very large n. And under almost none of those options will *any* of p1 – p5 exist ever at all. This way of looking at things shows why Parfit and others have taken it that, looking at things from the perspective of p1 – p5, our staying the course, our choice of S, was in fact their best option. The p-persons have not been made worse off by what we have done; they have not been harmed by what we have done; and they have incurred no loss whatsoever as an effect of what we have done. If S is wrong at all, it is wrong in virtue of something other than the fact that it makes things worse for persons who do or will exist under S.

Now, this is not to say that there are not *some* ways of realizing the "alter the course" option that *do* produce at least some p-person. Witness A1, which, performed at  $w_{A1}$ , generates an outcome that includes each of p1 – p4 and makes them all better off. (A1 produces p5 as well; but any plausible tradeoff principle would say the tradeoff is well worth making.) We don't delete A1 from the graph. We can't plausibly argue that A1 isn't in the feasible set—or that its outcome for the p-people, described at  $w_{A1}$ , isn't a perfectly accessible way in which the future might itself unfold. Climate change is not, in other words, a "*can't-do-better*" problem. Instead, it is a very-very-probably-can't-expect-better-than-what-has-been-done" problem—it's a "*can't-expect-better*" problem, for short.

But that fact does not (or at least, not immediately) allow us to discern harm to any p-person. No one is denying that the p-people are better off under A1 than they are under S. The point is, rather, that there's no reason to think that, if agents refrain from performing S, they'll land on A1 instead. The probability is overwhelming that they will instead land on a member of that very large collection of alternative ways of realizing choice A under which no p-person ever exists at all.

### 3. *The Fallacy of the "Can't-Expect-Better" Problem*

Still, there's something odd going on here. We are comparing, for each p-person, the probability of that person's coming into existence if we choose A in place of S, against that person's actual wellbeing level, given that we do perform S. But that's just to compare a number between 0 and 1—that is, a probability—against an actual individual wellbeing level. That the latter is greater than the former doesn't tell us anything about whether the latter is better for p1—say—than the former is. It's like saying that \$5 being handed to me in cash is better for me than, oh, a probability of 0.1 "being handed to me." Despite the fact that we can easily say that 5 is *greater than* 0.1, such comparisons aren't well-defined. If it's a 1 in 10 chance of winning a billion dollars, then the fact that 5 is greater than 0.1 doesn't even begin to show that being handed \$5 is better for me than being handed the 1 in 10 chance of winning a billion dollars.

Now, to this quibble, the can't-expect-better problem has a ready answer. It's easy enough to explain why the very small probability of any p-person coming into existence under our choice of A is important in this context. We aren't here really comparing a number between 0 and 1 that represents the probability of any p-person's coming into existence, given A, against that person's actual wellbeing level, given S. The probability of any such p's coming into existence is instead relevant by virtue of the role it plays in our calculation of the *expected value* for any such p of the choice of A ( $EV(A, p)$ ). And here we calculate expected value in the usual way: we take the sum of the value of each outcome for each p-person (one-by-one), multiplied by the probability that that outcome will obtain, given our choice of A. Of course, the *value* of the outcome described in  $w_{A1}$  for—say—p1 is nice and high—it's +7, by supposition; but the *probability* that that outcome will obtain for p1 is very, very low. The far more likely outcome is the one we see in  $w_{A2}$ ,  $w_{A3}$ , etc.: the outcome where p1 never exists at all. What we include in the summation for those cases is just zero—and that's so, taking it for granted for the moment that the value of never existing at all for p1 is just zero.

The upshot is:  $EV(A, p_1)$ —of not doing precisely what we did; of not performing  $S$ —is itself very, very low: it's some very large sequence of zeros, preceded by a decimal point, and followed by a 7. It's just barely there at all. The flawed existence we see for  $p_1$  under  $A$  is much greater than that: a nice, fulsome  $\approx 3$ . And so we conclude as before.

There's still something odd going on here. We are now overtly comparing the expected value of the choice of  $A$  against the actual value of  $S$ . But that is not going to be a valid way of determining betterness, or calculating harm, or what I have here called loss. There is an issue whether we should take an actual value approach across the board, as Fred Feldman does, or an expected value approach across the board, as Peter Singer does, and there are arguments to be made on both sides of that issue. But there is no issue whether we can take a hybrid approach. We can't. For too many acts  $X$  and  $Y$  and persons  $p$ , the  $AV(X, p) > EV(Y, p)$  and the  $AV(Y, p) > EV(X, p)$ . But we can't say that  $X$  is both better than  $Y$  for  $p$  *and* that  $Y$  is better than  $X$  for  $p$ , without *completely abandoning* the logical features that we need to attribute to betterness-for- $p$  to produce a relation we understand and make use of for determining whether a harm, or loss, has been imposed or not.

If we take an actual value approach to the question of betterness, and harm or loss, then we are done. The fallacy of the nonidentity problem is the comparison it urges us to make between (1) the actual value of how the future has unfolded for  $p_1$  under  $S$ , and (2) the expected value of  $A$  for  $p_1$ ; and certainly the former number is greater than the latter. But that's no indication that that  $S$  is better for  $p_1$  than  $A$  is—or, more to the point, that  $S$  does not harm, or impose a loss, on  $p_1$ .

But at this juncture the nonidentity advocate is going to insist—as Parfit himself does, in *Reasons and Persons*—that our approach should be prospective in nature. And that's to say (I take it) that we should be comparing the expected value of  $A$  for  $p_1$  against the expected value of  $S$  for  $p_1$ .

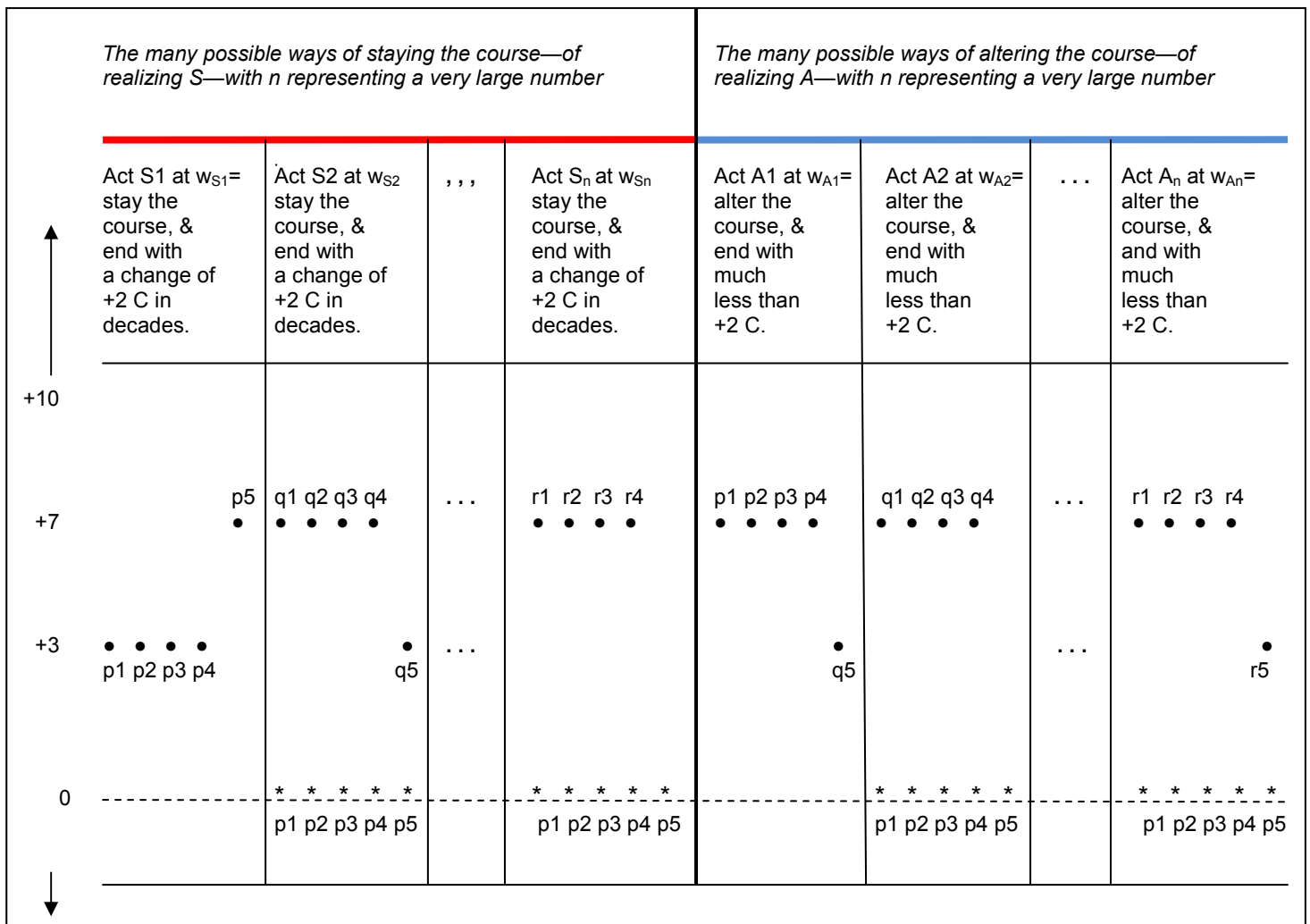
We've conceded that  $EV(A, p_1)$  is very, very low. But what is the  $EV(S, p_1)$ ?

Now, to figure that out, it's a mistake to attribute to agents information that they are only in a position to grasp once the future has in fact unfolded in one way rather than another. The whole point of expected value theory is to give agents a means, at least in principle, of evaluating their alternatives *prior to choice*. It is very hard, once we know something—here, how the future has in fact unfolded—to mentally go back into time, and to calculate *as though* we did not know that thing at all. But that's just what we must do, in order to calculate EV (S, p1) in an accurate way.

What we know *after the fact* is that agents realized S in a way that ends in the coming into existence, decades, or centuries, later, of p1 – p5. But before the fact, agents had no inkling that they would realize S in exactly the way that they did. Just as there are a zillion alternative ways of realizing A, so are there a zillion alternative ways of realizing S. Just as the agents, in advance of choice, have no way of realizing A in a way that insures p1's coming into existence, so do they have no way of realizing S in a way that insures p1's coming into existence. Just as the probability of p1's coming into existence, given A and given the world as it existed just prior to choice, is practically nil, so is p1's coming into existence, given S and given the world as it existed just prior to choice.

In fact, given the random hazards that shall surely come with climate change, isn't it in fact lower? There is an idea floating around, courtesy of the nonidentity problem, that children born of Holocaust survivors, in the late forties, or fifties, or sixties or even more recently, "owed their very existence" to the fact of the Holocaust. Those individuals may have been raised with no grandparents, no extended family; and their family may have lost all of their hard-earned assets that otherwise would have been handed down from one generation to the next; but still they personally, despite the adversity, incurred no loss as a result of the Holocaust, but rather, if anything, a gain. I think that's a mistake. Any such person's coming into existence was highly improbable in any event, at that moment just prior to choice; but any such person's coming into existence was surely made still more improbable as a result of the Holocaust.<sup>3</sup>

The accurate presentation, then, is no more Graph II than it is Graph I. What we need is a third graph—a graph that recognizes that the point about the precariousness of any particular person's coming into existence cuts both ways, and that as to the probability of any particular person's coming into existence, the choice to stay the course and the choice to alter the course are at best a *wash*.



GRAPH III: CLIMATE CHANGE + NONIDENTITY—CORRECTED TO SHOW THAT THE PRECARIOUSNESS OF EXISTENCE CUTS BOTH WAYS AND THAT ANY GIVEN PERSON'S COMING INTO EXISTENCE IS HIGHLY IMPROBABLE UNDER BOTH S AND A.

Now, we may well think that the expected value against expected value comparison is pertinent to determining whether one act is better for a person than a second, but that one act is "prospectively better" than another is perfectly consistent with that person's not incurring any loss at all as a result of the performance of that second act. We may well think that an elevated risk is not itself a loss or a harm. I am attracted to that view myself: misses don't count. But climate change isn't a miss: at least at the moment, we have no basis whatsoever for concluding that we are here dealing with a risk that will never eventuate. Instead, we have every reason to suppose that, if the risk is imposed, it will eventuate—that the picture we have drawn under S1 will not be accurate. It's always possible that somehow climate change ends up, not ruining the environment, but rescuing it. But we have no basis whatsoever for accepting that particular prediction at the present time—and every reason to reject it.

#### 4. *Does this Person-Affecting Approach Commit Us to Moral Actualism?*

We've assumed that S is what is actually done, and that S has been realized by S1 and that p1-p5 actually exist as a result. What grounds our result that S is wrong is:

- the availability of A1 as an alternative to agents at the critical time
- the accessibility of  $w_{A1}$  at the critical time
- the actual value of A1 being greater for p1-p4 than the actual value of S1
- the fact that the probability, *as calculated on the basis of the information that is within the agents' grasp prior to choice*, of any person's, including any p-person's, coming into existence, given A, is, while very, very low, no less than the probability of any such person's coming into existence, given S
- the correlated fact that the expected value of A for any p-person is less than the expected value of S for that same person

--and finally, the fact that securing the better result for p1-p4 does not impose too deep a loss on anyone else who does or will exist under A1, that is, on p5).

What I want to clarify now is that my proposed analysis does not include any implication that agents are *obligated* to perform A1—as opposed to A2, A3 and so on.

Why might one think anyone else? According to my proposed analysis, if the reason S1 is wrong is what S1 does *to the actual persons p1-p4*, doesn't that mean that p1-p4 matter morally in some special way? And if that is so, then why aren't A2, A3 and so on "just as wrong" as S1 is? In fact, are they not still more clearly wrong, in view of the fact that A2, A3 and so on impose a significantly graver loss on p1-p4 than S1 does?

Such results are exactly what we would obtain from the highly problematic *moral actualism*, according to which actual persons and only actual persons, and their losses and only their losses, matter morally. But my proposed analysis does not need, and shouldn't be taken to endorse, that principle or any of the others that have sometimes been associated with the person-affecting approach and have been effectively shown to be problematic.<sup>4</sup>

I grant that, just as p1 incurs a loss under S1, p1 incurs a loss under A2, and in fact incurs a deeper loss. But we can say that the loss p incurs under S1 *counts against* S1—and that, given the alternatives available to the agents at the critical time, *it makes S1 wrong*—and still say that p1's loss under A2 does not count against A2 at all. For the position that *one* of p1's losses—the loss p1 incurs under S1—is morally significant and counts against S1 does not even begin to imply that *all* p1's losses are morally significant—that they *all* count against the particular acts that impose those losses and in a roundabout way in favor of the acts that avoid those losses. We can, and should, instead say that some of p1's losses are morally significant and some are not, with the moral significance of any given loss being a function of just *where* that loss is incurred in relation to the person who incurs it. That is: incurred at a world,  $w_{A2}$ , where p1 never exists at all, p1's loss is devoid of moral significance; and, incurred at a world,  $w_{S1}$ , where p1 does or will exist, p1's loss has

Roberts, "Counting the Costs Before They're Hatched"

full moral significance. According to this principle, which we can call *Variabilism*, p1, like each other person, whether actual or not, matters morally but matters *variably*.<sup>5</sup> And we can then explain that, while S1 is wrong, A1 itself is not obligatory but rather permissible, given A2, A3 and so on: it is important to understand that each loss incurred by each p-person under S1 has full moral significance, but just as important to understand that the loss we impose on each p-person if we happen to perform one of A2, A3 and so on, in place of A1, is devoid of any moral significance whatsoever.

---

<sup>1</sup> John Broome seems to accept this argument. See *Counting the Costs of Global Warming* (Cambridge: White Horse, 1992). See generally Derek Parfit, *Reasons and Persons* (Oxford: 1987), pp. 351-79.

<sup>2</sup> I am thinking here of the repugnant conclusion; the infinite population problem; the equality problem; and the Asymmetry. But see Broome's discussion of the repugnant conclusion in *Weighing Lives* (Oxford University Press, 2004).

<sup>3</sup> In connection with Parfit's Depletion Example—which is also a can't-expect-better problem—it was recently pointed out to me by Peter Singer that—given the expectation of the less than healthy future environment—p's coming into existence may well be *less likely* under the choice of Depletion than it is under the choice of Conservation.

<sup>4</sup> This disavowal of the moral actualism includes moral actualism in both its truly actualist, or "strong," and "weak" forms. See Caspar Hare, "Voices from Another World," *Ethics* 117 (2007), pp. 498-523. Nor am I adopting what Peter Singer calls the "prior existence view," or what Broome calls the "neutrality intuition" and develops as the "principle of equal existence." See Singer, *Practical Ethics*, 2<sup>nd</sup> ed. (Cambridge 1999), p. 103; Broome, *Weighing Lives*, pp. 143-146. I am in agreement with Hare, Singer and Broome's that each of these principles is untenable.

<sup>5</sup> I am very grateful to Saul Smilansky and Mark Greene for their recent discussions with me of the can't-expect-better problem. I have explored this problem elsewhere. See Roberts, "The Nonidentity Fallacy: Harm, Probability and Another Look at Parfit's Depletion Example," *Utilitas* 9: 267-311(2007), and "The Nonidentity Problem and the Two Envelope Problem: When Is One Act Better for a Person than Another?" in *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*, eds. M. A. Roberts and D.T. Wasserman (Springer, 2009), pp. 201-228.