CANADIAN JOURNAL OF PHILOSOPHY, 2017
VOL. 47, NOS. 2–3, 200–228
http://dx.doi.org/10.1080/00455091.2017.1279518

Routledge
Taylor & Francis Group

Is a person-affecting solution to the nonidentity problem impossible? Axiology, accessibility and additional people

Melinda A. Roberts

The Department of Philosophy, Religion and Classical Studies, The College of New Jersey, Ewing, NJ, USA

ABSTRACT

This paper considers two objections based in axiological considerations against the position that whether a given outcome, or possible future or *world*, is *morally worse* than a second world may depend in part on what is going on at a third world. Such a wide-angled approach to determining worseness is critical to the solution I have previously proposed in connection with the nonidentity problem. I argue that both objections fail.

ARTICLE HISTORY Received 3 January 2017; Accepted 4 January 2017

KEYWORDS Nonidentity problem; axiology; accessibility; person-affecting intuition

1. Introduction

This paper has two main goals. The first is to articulate an objection against the person-affecting solution to the nonidentity problem that I have proposed elsewhere.[1] Critical to my solution is the claim that whether a given outcome, or possible future or *world*, is all-things-considered morally worse than a second world may depend in part on what is going on at a third world. The objection is that axiological considerations show that that sort of wide-angle-lens approach to determining worseness between worlds leads to inconsistency. The second goal of this paper, then, is to argue that that objection fails.

Still a third goal of this paper is to evaluate an ancillary objection. That objection is based on the claim that any approach that ties outcome worseness to a notion of when one world is *accessible* to another – which mine does – must rely on an overly liberal notion of *accessibility* in order to satisfy still other axiological constraints. I argue, however, that a liberal notion of accessibility is not obviously

CONTACT Melinda A. Roberts robertsm@tcnj.edu

© 2017 Canadian Journal of Philosophy

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

problematic and that, indeed, it might turn out to be useful in addressing not just the nonidentity problem but other problems in ethics as well.

I begin the paper, in Section 2 below, by noting that Rawls' *separateness of persons* objection to classical utilitarianism strongly motivates what is called the *person-affecting intuition*. But – at least in connection with the *telic* component of that intuition – formulation problems immediately arise. Standardly formulated, the core intuition – which is itself simply a necessary condition on when one world *w* is all-things-considered morally worse than another world *w*′ – allows much to turn on an *unduly narrow inquiry*. The standard formulation forces us in many cases to conclude that *w* isn't morally worse than *w*′ without *any consideration* of how things in some third available, or (I'll say) *accessible*, world *w*″ turn out for the very person who at first glance looks to have been disadvantaged by how things have in fact turned out for that person in *w* and who does not exist at all in *w*′.

Having truncated the discussion, the standard formulation effectively *rules out* what would otherwise constitute a highly plausible person-affecting solution to the version of the nonidentity problem that I consider the most challenging.

On the other hand, the standard formulation of the core person-affecting intuition is hardly the only way of formulating that intuition. Thus, as I also argue in Section 2 below the better candidate for that job comes in the form of an *alternate* formulation – a formulation that insists on a more *far-reaching inquiry* before generating the result that *w* isn't worse than *w*′.

I then briefly sketch, in Section 3 below, the role that that wide-angled, alternate formulation of the intuition plays in analyzing the nonidentity problem. I'll focus on Kavka's *pleasure pill* case. But it will be obvious that that same analysis can be extended to many other nonidentity cases as well, including Kavka's slave child case, Parfit's depletion and risky policy examples, cases involving historical injustices and, finally, cases involving environmental malfeasance, including do-nothing policies on climate change.

In Section 4 below, I worry. As noted, the core person-affecting intuition is standardly formulated in terms that allow much to turn on an *unduly narrow inquiry*. Philosophers presumably have their reasons for formulating the intuition in that sure-to-fail way. I try to identify what those reasons might be. The most potent objection against the alternate formulation that I favor – an objection that tacitly approves the standard formulation – is that axiological considerations show that that alternate formulation leads to inconsistency. I argue, however, also in Section 4, that that objection fails. Other philosophers have attempted to circumvent the inconsistency objection. Thus, Temkin, for example, proposes that the defect in the argument is that it assumes that worseness between worlds is *transitive*.[2] But I believe that the objection instead fails in virtue of the fact that it fails to deploy resources that we are allowed by the very theory the objection is meant to refute.

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

I also in Section 4 briefly address an ancillary objection to the alternate formulation. According to that objection, an approach that ties outcome worseness to a notion of when one world is *accessible* to another must involve a notion of *accessibility* that is itself overly liberal. I concede that the notion of *accessibility* that I make use of may seem to count against the convenient position that an agent's act must be deemed permissible if that act ends in an outcome that represents the best that that agent *working on his or her own* has the ability, the resources and the power to bring about. But I argue that a liberal notion of accessibility isn't itself obviously problematic and, indeed, may turn out to be useful in solving not just the nonidentity problem but other problems in ethics as well.

I conclude, in Section 5, with a note on a variation on Parfit's *mere addition paradox*, that is, *addition plus*. I point out that my wide-angled, alternate formulation of the person-affecting intuition produces desirable results in that case. And I argue that failing to deploy the complete range of resources that come with our own theory can quickly lead us to think that such puzzle cases are even *more* intractable than they already seem to us to be. Addition plus, like mere addition, will never be easy, but there's no point making it harder than it is.

2. The person-affecting intuition

2.1. The separateness of persons objection to totalism

Classical aggregative utilitarianism (*totalism*) was said by Rawls to fail to take into account the 'distinction between persons.'[3] Exactly what the separateness objection comes to and whether it is probative have been much discussed in recent years.[4] We might, however, consider Rawls' point not a definitive objection against totalism but rather a rough diagnosis that directs our attention to the *mechanism* by which totalism produces results we think are clearly false in – among others – simple equality and additional person cases.

According to that diagnosis, the problem with totalism is that it evaluates a given world by simply *aggregating* across the population *whatever it is* – happiness, preference satisfaction, capability, resources, whatever – that gives existence its value for each person who does or will exist at that world. I will call that thing *well-being* and understand that for a person to have more *well-being* in one world than another is just for the one world to be *better for* that person than the other. Then, according to totalism, for one world to be *better* than another is just for the one world to contain *more well-being in the aggregate* than the other does.

It's true that, under that principle, a reduction in any one person's well-being level will *automatically* mean (other things equal) a reduction in aggregate well-being – in, that is, *overall* value. For that reason, totalism can't credibly be accused of ignoring the plights of anyone at all. But perhaps what the

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

'separateness of person' objection is meant to point out is that totalism doesn't take those plights into account *in the right way*. For example, it doesn't put limits on how far a *particular* person's well-being might fall as long as, inch by inch, that fall is offset by gains in well-being that might themselves be thinly spread across a sizeable number of people who will remain very well off *however* the future unfolds. Nor does it put limits on how far a given person's well-being might fall as long as, inch by inch, that fall is offset by gains in well-being accrued by *additional* people, that is, people who need *never have existed at all*.

Consistent, though, with what we've just said is that the problem with totalism isn't that it *aggregates* well-being and thus fails to distinguish one *person* from another but rather that it aggregates *well-being* and thus fails to distinguish one sort of *loss* from another. Thus, let's imagine that losses of (or gains in) well-being come in *units*. We might accept that an additional unit of well-being makes things no less better for a better off person than that additional unit makes things for the less well off person. We might accept that the one unit has the same *personal* value, whoever happens to get it. *But* consistent with that we might also accept that the unit that makes things better for the better off person may have no *moral* value at all. It may not make things *morally* better. But allocated to the *less* well off person – well, *that's* a different story. That one unit may then have *full* moral significance; it may then *easily* make things morally better.[5]

Ditto when the trade-offs are between an *existing* person who will either incur or not incur a loss of well-being and a *possible* person who will either incur or not incur a loss of well-being as a function of whether he or she happens ever to exist. Perhaps the lost unit of well-being matters morally when the loss is due to how the world unfolds for the *existing* person but not when it's due to the *possible* person's never existing at all. Perhaps, correspondingly, the gain in well-being that makes things better for the *possible* person by way of bringing that person into existence *doesn't* make things morally better, whereas the gain in well-being that makes things better for the *existing* person *does* make things morally better.[6]

Rawls' comment thus can be subjected to at least two distinct analyses. Whichever we choose, however, the conclusion may be the same. The problem with totalism is that it *aggregates well-being*. Either we should *not aggregate* to begin with or, if we do aggregate, we should aggregate *something other than well-being*.[7]

2.2. Some cases

Consider the *simple equality case*. Agents have exactly two options. They can produce a population of 99 people *p1 … p99* who enjoy unimaginably wonderful lives along with one person *q* who has a life of complete squalor, degradation, pain, misery and suffering. Whatever we think it takes to bring it about that *q*'s life is *far* less than worth living, that's what we should imagine happening to *q*.

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

Alternatively, agents can produce exactly the same population of individuals, each of whom, including *q*, enjoys a life that is *just short* of the unimaginably wonderful.

Simple equality case		
Well-being	*w1* (including *a1*)	*w2* (including *a2*)
+100	**p1 … p99**	
+98		**p1 … p99, q**
…		
−100	**q**	

Totalism declares that the first option, or outcome or possible future or world, is just as good as the second. We agree that that's a mistake. It's a mistake to deem the loss to *q* in *w1* to be completely offset – whitewashed away – by gains to *p1* … *p99*. It's a mistake not to accord to *q*'s very deep loss in *w1* a *moral significance* that the relatively shallow losses *p1* … *p99* suffer in *w2* lack, and it's a mistake not to recognize that *q*'s gain in *w2* has a *moral significance* that the gains *p1* … *p99* enjoy in *w1* lack. But the offsetting itself seems just a function of the particular aggregative principle that totalism embraces. Hence, the diagnosis: totalism produces results that we think are clearly false in virtue of the fact that it aggregates well-being across the population. So, we shouldn't aggregate well-being across the population.

That totalism is oblivious to the distinctions we have just made is one strike against totalism. Perhaps, of course, the problem with totalism is not that it's impervious to the value of *equality*; perhaps its real failing is that it's impervious to values that prioritarianism, or perhaps leximin, are designed to capture. Whatever are the finer-grained features of our diagnosis of the problem with totalism, the account it offers of the simple inequality case should make one point clear: that *more aggregate well-being* doesn't necessarily make for a *morally better* world.

Totalism produces results that seem just as clearly false in the *additional person case* – and it seems to generate those false results using the same defective mechanism that was at work in the simply equality case. Here, boldface indicates a person who does or will exist, and italics paired with the asterisk a person who never exists, at the indicated world.

Additional person case		
Well-being	*w1* (including *a1*)	*w2* (including *a2*)
+10	**q**	**p**
…		
+0	*p**	**q**

Totalism concludes that *w1* and *w2* are equally good – that neither world is better than the other, and that the acts that end in those different worlds are both permissible – by way of simply aggregating well-being across the populations at *w1* and *w2*. But surely we can agree that *w1* is better. Totalism fails because it ignores the distinction between *q*'s loss in *w2* and *p*'s loss in *w1*. It leaves us with no room to say that *q*'s loss in *w2* is morally significant in a way that *p*'s loss in *w1* isn't. But that seems exactly what we want and need in the end to say.

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

2.3. The person-affecting intuition as antidote to totalism

Let's suppose, going forward, that either we should *not aggregate* to begin with or, if we do aggregate, we should aggregate *something other than well-being*. That we should reject totalism doesn't mean that we should reject the many features totalism has that we find compelling, including its insistence that agents maximize the good and do the best they can for people and that acts are to be evaluated on the basis of their consequences.[8]

But can a theory retain those features but determine whether one world is morally worse than another *without* aggregating well-being across the population?

Perhaps it can, by shifting from talk about maximizing well-being *in the aggregate* to talk about maximizing well-being *for each person as an individual*.[9] Thus, a *person-affecting* account of the additional person case might start with the core principle that, where the well-being of *each and every person who does or will exist* at a given world has been maximized, nothing at that world counts against what has been done there or makes the world itself morally worse than any other world. In other words:

> P*: For each world and each person who *does or will exist* at that world, if well-being is *maximized* for that person at that world, then nothing *counts against* acts performed at that world and nothing makes that world *worse than* any other world.

A plausible, person-affecting account of the simple equality case would need to say more since such an account must address how conflicts are to be resolved and trade-offs made. But the underlying goal of the person-affecting intuition – and of the person-affecting approach more generally – for both the additional person case and the simple equality case would be to attend carefully in a way that totalism does not – attend, that is, in the *right way* – to whether a particular loss – *q*'s loss in *w1* in the simple equality case; *q*'s loss in *w2* in the additional person case – may have a moral significance that other losses lack.

It's the additional person case that we need to focus on for purposes here. A few notes are in order. First, P* is restricted to people who *do or will exist* at the world under evaluation. I believe that that restriction is critical. For the claim that *q*'s loss in *w2* is just as deep as *p*'s loss in *w1* – that *w2* is worse for *q* by the same amount that *w1* is worse for *p* – seems to me both plausible and, though controversial, hard clearly to refute. If that claim stands, and if we *don't* restrict P* to people who do or will exist, then P* cannot help us make the distinction between *q*'s plight in *w2* and *p*'s in *w1*. But the whole point – or at least, a big chunk of the point – of the person-affecting approach is to make exactly that distinction.[10]

Second, P* provides only a *necessary* condition on when an act is morally wrong and a world morally worse than another. P* isn't, that is, *definitional*.

And finally one addition. We can now just add that well-being is *maximized* at a world *w* for a person *p if and only if* there is no other world *w'* available to

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

agents from *w*, that is, *accessible* to *w*, such that *p*'s well-being in *w*′ is greater than *p*'s well-being in *w*.[11]

For purposes here, it's useful to separate out P*'s *deontic*, or act-evaluating, component, and its *telic*, or outcome-evaluating, component.

Taking the contrapositive of P* and borrowing from Parfit, we can put the deontic component this way:

> P* *(deontic)*: What is 'bad' (morally wrong) must be 'bad for' (make things *worse for*) a person *who does or will exist* at the world where the act under scrutiny is performed.[12]

The telic component, in contrast, is rarely so well-constructed. Thus, it's often formulated elliptically, as follows: for one world to be worse than another, the one world must be *worse for* a person who does or will exist at that world.[13] But *worse for* a person who does or will exist at that world *than what*?

P* itself provides an answer to that question. Before licensing the critical inference, the wide-angled P* requires a more far-reaching inquiry, an inquiry not just into how a person *p* fares in *w* as compared against *w*′ but against all other accessible worlds *w*″ as well. P*, in effect, makes the test for when one world is worse than another harder to fail. Thus, again taking the contrapositive, we can write:

> P* (*telic*): A world *w* that is worse than a world *w*′ must be worse for a person who does or will exist at *w* than a world *w*″ accessible to *w* is for that person, where *w*″ may but need not be identical to *w*′.

I believe that the ellipsis must be filled out in just this far-reaching way if the resulting principle is to stand up to the nonidentity problem – or indeed to have any chance of acceptance at all.

It's thus worrisome to me that *no one does it this way*. Rather, in cases where the ellipsis is filled out at all, it's filled out in a way that forces us to conclude that *w* isn't worse than *w*′ on the basis of a *very narrow inquiry*. Specifically, the industry standard for filling out the ellipsis is to require that *w*″ be identical to *w*′ – to, that is, in effect omit *w*″ from the picture altogether.

But where worlds *w* and *w*′ are the worlds that we are to rank, and the case is one in which the person who seems to have been disadvantaged exists in *w* but not in *w*′, that way of filling out the ellipsis forces us to conclude that *w* is at least as good as *w*′ is *unless* we find that *w* is actually *worse* for that person than *never existing at all* would have been.

But that's a pretty easy way of jacking up the overall good of *w*. In fact it's *too* easy. How a person *p* is treated in *w* may be an absolute *travesty* – compared, not against how *p* is treated in *w*′, but against how *p* is treated in *w*″. Why shouldn't that latter fact count *against w*? Why should that latter fact be deemed immaterial to the analysis? Why should the person-affecting theorist accept a requirement that dictates that a person-affecting comparison between *w* and *w*′ can be completed *without reference to what is going on in any third*

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

world w″ even in a case where what is going on in *w″* seems *painfully relevant* to any accurate evaluation of *w*?[14]

That philosophers standardly opt for such a sure-to-fail formulation of the telic component of the person-affecting intuition – a principle that, as we shall see, bars a plausible person-affecting solution to the nonidentity problem and a principle that on other grounds we will be forced to reject – in place of what P* itself suggests of course gives us pause. And pause we shall.

But first some of the plusses of P* should be at least noted. For example, P* opens the door to an entirely plausible account of the additional person case. Thus, P* instructs that *w1* is without any moral deficiency – that is, that the choice, or act, *a1* is permissible, and *w1* isn't worse than *w2*.

Now, P*, as noted, provides only a *necessary* condition for when an act is wrong and a world worse. But P* is perfectly consistent with further person-affecting principles, including certain *sufficient* conditions on wrongdoing and moral worseness, that seem entirely uncontroversial. Such principles can be expected to tell us, among other things, that what happens to the existing *q* at *w2 does* count against what is done at *w2* and *does* make *w2* worse than *w1* is.[15]

We can put these points another way. P* gives us the resources we need to say that *p*'s loss at *w1* just doesn't count – and also gives us the room we need to say, now putting still other person-affecting principles to work, that *q*'s loss at *w2 does* count. And these are just the things that we want to say about the additional person case.

2.4. The moral status of merely possible people

The fact that P* makes moral worseness, and wrongdoing, a matter of how people who *do or will exist* fare at the world in question may seem to suggest that accepting P* requires a commitment to the position that *existing* and *future* people have some sort of a *moral status* that the *merely possible* lack.

Thus, it may seem that anyone who accepts P* must surely accept the Pareto-like Px as well.[16]

> Px: For each world *w*, if there is a world *w*′ accessible to agents that is better for a person who *does or will exist* in *w* and is not worse for any person who *does or will exist* in *w*, then what agents have done at *w* is wrong and *w* is worse than *w*′.

It's worth pointing out that Px actually does a *better* job with some additional person cases than certain other principles that have been explored in connection with the person-affecting approach do. I am thinking here of the *prior existence view* (PEV).[17]

Consider, for example, the *procreative asymmetry*. According to that two-part intuition, it is (other things equal) wrong to bring a *miserable* child – a child whose life is *less* than worth living; a child whose well-being level falls into the negative range – into existence but permissible *not* to bring a *happy* child into

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

existence. The second half of that asymmetry aligns nicely with the account P* suggests for the additional person case. But it's the first half, the miserable child half, of the asymmetry we want to focus on here.

According to Px, if it would have been better for the miserable child for that child never to have existed at all, and if leaving that child out of existence would not have made things worse for any other existing or future person, then it's wrong, and makes the world worse, to bring that child into existence. That result seems right.[18]

Consider, in contrast, the account PEV offers of the miserable child case. According to PEV, we 'should promote the welfare of existing beings and those who will exist, independently of our decisions.'[19] PEV thus finds 'no grounds for objecting to' the decision to bring the miserable child into existence.[20] And that result seems clearly false.

Of course, that Px does a better job with the miserable child half of the asymmetry than PEV does doesn't mean that Px is true. Px is susceptible to its own counterexamples and must be set aside as well.

Consider the following case. Let *n* be *very* large.

Merely possible people matter case		
Well-being	*w1* (including *a1*)	*w2* (including *a2*)
+10		**p**
+9	**p**	
…		
+0	*q1* … qn**	
−1000		**q1 … qn**

One might even change the example by eliminating '*qn*' and letting the class of *q*-persons be *infinitely* large. Either way, Px tells us that *w1* is worse than *w2*. But that result seems clearly false, and so we reject Px.

There's a larger lesson here. It won't do to oversimplify our discussions of obligations in respect of the merely possible. It might seem tempting to say that any third child I might have produced in addition to Annabel and Tommy must be, as a never-existing or indeed non-actual being, devoid of moral status – and that *that's* why it was permissible for me *not* to produce such a third child. But that temptation should be resisted. If my not producing a third child was indeed permissible, it's not because that additional, never-existing, non-actual child has *no moral status*. Rather, something else is going on, something that holds me accountable for many of the losses Annabel and Tommy might incur but *not* for the very deep loss incurred by any such third child. We shall return to what that something else might be in Section 2.5 below. But for the moment the important thing is to underline that it won't do to connect *moral* status to *modal* status – that is, to say that some people (actual people, or perhaps existing people) have full moral status while other people (non-actual, never-existing people) have none at all. Px is false.

Another problem case for Px is *addition plus*, a variation on Parfit's *mere addition paradox*.

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

Addition plus			
Well-being	*w1* (including *a1*)	*w2* (including *a2*)	*w3* (including *a3*)
+10		**p**	
+9	**p**		
+5			**p, q**
+1		**p**	
+0	*q**		

Px implies that *a1* is wrong and that *w1* is morally worse than *w2* is. The better view, however, is that the plight the never existing *q* faces in the non-actual world *w2 counts against* what is done at *w2* and makes *w2*, not just worse than *w3*, but – I believe – worse than *w1* as well. What is uncontroversial is that what happens to *q* in *w2* makes *w2* worse than *w3* is. After all, *w3* makes things better for *q* than *w2* without making things *too* much worse for *p*. What is perhaps controversial – though I think correct – is that what happens to *q* in *w2* also makes *w2* worse than *w1* is *even though w1* is worse for *q* than *w2* is. What happens to *q* in *w2*, in other words, counts not just against *w2* as compared against *w3* but also against *w2* as compared against *w1*. But that means that *q*, after all, matters morally, *even though q* is *merely possible* relative to *w1*, for purposes of both evaluating *a1* and comparing *w1* against *w2*.

Such an account of addition plus suggests, then, the following: never-existing, non-actual merely possible people matter morally, just as you and I matter morally, and they matter *radically*: they matter morally, not only at the never-existing and non-actual worlds where it so happens that they do or will *exist*, but also at worlds where they *never* exist at all.

2.5. The moral status of the merely possible; implications for P*; variabilism

We can itemize some general points regarding the moral status of possible people, a class that includes you and me and people who don't now but will at some future time exist and people who don't now and will never exist.

(a) *Each person has the very same moral status as each other person*. People who are, relative to a given world, *merely possible* matter morally in just the same way that people who do or will exist in that world matter morally.

(b) *Actual people have no special moral status*. It's not the case that *actual* people, that is, people who do or will exist at the uniquely actual world, have a moral status that others lack. And it's not the case that the moral status actual people have is greater than the moral status other people have.

(c) *Existing (and future) people have no special moral status*. It's not the case that the only people who matter morally for purposes of evaluating a given act at a given world are the people who do or will exist under that act and at that world. In addition plus, *q* matters morally for purposes of evaluating *a1* even though *q* never exists under *a1*.

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

(d) *Moral status not 'conditional' on existence*. It's not the case that the moral status of a given person is *conditional* on that person's existence. Again, in addition plus, *q* has moral status relative to *w1* and *a1* even though *q* never exists at *w1*.

(e) *Moral status necessary*. We take for granted that people have moral status in worlds where they do or will exist. Bad things that happen to them there count against what is done there and shift the balance in the overall comparison of one world against another. But there's more: the person who does or will exist in one world has moral status in *both* that world *and* in every other world. The bad thing that happens to a person in a world where that person does or will exist *bears on* the evaluation of what is done at worlds where that person never exists.

(f) *No discounting* of the losses incurred or the gains accrued by the merely possible. It's just as important that the lives of the merely possible go well at a *non-actual* world where those people have lives as it is that our lives go well at the *actual* world. Thus, in addition plus, if improving *q*'s plight in *w2* counted for only 10% of what improving *p*'s plight in *w1* counts for, then agents are obligated to improve *p*'s plight. But we reject that result.

The foregoing compendium of claims regarding the moral status of the merely possible might well lead one to think that we must reject P* – along with the happy child half of the asymmetry and, in the additional person case, the idea that nothing at *w1* counts against what has been done at *w1* or against *w1* itself. If (1) the merely possible matter morally in exactly the same way you and I matter morally, and if (2) at no cost to anyone else bringing a merely possible person into an existence worth having makes things better for that person, then doesn't it *follow* that (3) *failing* to bring that person into existence (other things equal) is wrong and makes the world worse and (where other things aren't equal) at least counts against that act and that world?

No. We can accept (1) and (2) but still insist on P*. There's space between premises and conclusion. Thus, we can accept that all people matter morally and in exactly the same way but that there are *limits* to how any given person matters morally. In respect of each of us, what's important, from the *moral* point of view, is that diminutions in our well-being at worlds where we do or will exist be avoided. And, in respect of each of us, what's *not* important, from the *moral* point of view, is that diminutions in our well-being at worlds where we *never exist at all* be avoided. In respect of each of us, in other words, some *losses* of well-being – we might even say some *harms* – have full moral significance while others have no moral significance at all. The losses a person incurs at worlds where that person does or will exist have full moral significance – they count against those worlds and against the acts that bring those worlds about and in favor of alternate worlds and alternate acts. In contrast, the losses a person incurs at worlds where that person never exists have no moral significance at all.

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

That view, which I have elsewhere called *variabilism*, is pretty simple.[21] It's not exactly common sense since we might not be able to rely on common sense to recognize that merely possible people have the same moral status that you and I have. But it has many commonsensical implications, including that the losses Annabel and Tommy incur as existing people matter morally – and that I, as an important agent in their lives, was obligated to work to avoid (both before *and after* their conception, the nonidentity problem, as we shall see in Section 3 below, notwithstanding). But the loss some third child I never produced at all incurs in virtue of his or her never existing at all simply doesn't show up on the moral radar, either here, where that child never exists, or in an alternate world, where that child exists and is happy. The *loss* here doesn't count against what was done here or in favor of what was done there; the *gain* there doesn't count in favor of what was done there or against what was done here.

Ditto for you and me. Our parents weren't obligated to produce us or anyone else (other things equal) to begin with. But given that they did produce us, they had *many* obligations (*we* think!) to work (both after *and before* we were conceived, nonidentity problem, again, notwithstanding) to make our lives go well.

A word on what is called the *complaint view* is in order. We might – though it's kind of a waste of energy in view of how things turned out – rail against the prospect of our never coming into existence at all. We might, that is, *complain* at *this* world about our nonexistence at an *alternate* world at which we never exist. But it doesn't *follow* from that that our complaint has any moral significance. Similarly, we might at this world *applaud* our parents (express 'thanks') for their having brought us into existence. But it doesn't *follow* that that applause has any moral significance. Variabilism takes advantage of that logical failure, that gap. According to variabilism, our applause – our thanks – has *moral* significance *only if* what we are applauding *reverses* a *morally significant* complaint; and our complaint has *moral* significance only if the loss complained about is a loss incurred at a world where we *do or will exist*.[22]

Applied, then, to addition plus, variabilism yields the following: *q*'s ground for a *complaint* at *w2* is based on *q*'s being *worse* off in *w2* than *q* is in *w3* – it's *not* based on *q*'s being worse off in *w2* than *q* is in *w1*. It follows, as a conceptual matter, that *q* has a ground for *applause* – for *thanks* – at *w3* that is based on *q*'s being *better* off at *w3* than *q* is in *w2*. But it *doesn't* follow that *q* has a ground for applause at *w3* that is based on *q*'s being better off in *w3* than *q* is in *w1*.

3. The nonidentity problem

3.1. Two types of nonidentity problems

The nonidentity problem purports to be a counterexample against P* (in both its deontic and its telic forms). I have argued elsewhere that the person-affecting approach is capable of solving the nonidentity problem – but only if we first

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

recognize that the problem comes in distinct *types*.[23] Cases that give rise to the nonidentity problem – whatever their type – have the following pattern: (a) we think the act is wrong and the world worse but (b) we just *can't see* how any person who does or will exist under that act or at that world has been made worse off. However, it seems that for a certain subset of nonidentity cases – the *Type I* cases – the line of reasoning that results in (b) is based on a fallacy. Type I cases include Kavka's pleasure pill and slave child cases, Parfit's depletion and risky policy cases, cases involving historical injustices and environmental malfeasance (including do-nothing policies on climate change) and many others. In all such cases, I argue, closer inspection shows that we can readily identify just how an existing or future person has been made worse off. If that argument is correct, then the Type I problem simply dissolves.[24]

Other cases – the *Type II* cases – resist that analysis. They leave us no room to dispute (b); there is no fallacy that we can simply identify and then avoid.[25] But on reflection it also seems to me that the Type II cases do not describe a choice that is clearly wrong or a world that is clearly worse.[26]

An example of a Type II case would be a case in which agents choose to bring into existence a child with a genetic or chromosomal disorder but whose life nonetheless will be unambiguously worth living. To insure that we are reacting properly to the case – that is, that our minds are clear of any presupposition that might prejudice the evaluation – it's *critical* that we assume two further points: that it's *certain* that bringing the one child into existence will place no other existing or future person at all at *any* risk of *any* well-being loss, and that it's *certain* that the one child's own well-being has been maximized (that is, agents had no ability whatsoever, including by acts they may have performed before the child is ever conceived, to have made things better for that child). In such cases, it's not clear to me that the act under scrutiny is wrong or the world worse. If that's correct then, for the Type II problem, (b) stands but (a) can be rejected.

3.2. The pleasure pill case

My particular concern in this paper, however, is not the Type II problem but rather the Type I problem. For it is my proposed solution to the Type I problem that is put in jeopardy by the issues I noted at the beginning of this paper. If an axiological constraint we are compelled to accept rules out P* as an acceptable formulation of the person-affecting intuition, then the sort of solution I have proposed for the Type I problem shall inevitably fail.

But first I need just to sketch how that solution is supposed to work. The *pleasure pill* case works well for that purpose. There, a couple, prior to conceiving a child, 'paus[e]' to take a pleasure pill; their choice immediately 'change[s]' – from *what*, we might ask – the identity of the child then conceived.[27] But it also damages the embryo or fetus in some way, causing the child to be born 'mildly

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

handicapped.'[28] As such, the child will not be as well off as any *nonidentical* child who might have come into existence instead had the couple not paused to take the pill.

Let's call the child in fact brought into existence *Andy* and call any arbitrary, better off child the couple might have conceived had they not paused to take the pleasure pill *Ruth*.[29]

Type I nonidentity problem; modally impoverished presentation		
Well-being	*w1* (including *a1*) what *does* happen (agents pause to take pleasure pill)	*w2* (including *a2*) what *would have* happened (agents don't pause)
+10		**Ruth**
+8	**Andy**	
+0	*Ruth**	*Andy**

Does the couple's choice to pause to take the pill make things worse for Andy? The argument is that it does not – that, had the couple not taken the pill, Andy would never have existed at all.[30] And it's surely better to exist impaired – provided the existence is worth having; and here it's part of the case that the existence *is* worth having – than it is never to have existed at all.

The story itself – not the *graph*, but the *story* – gives us enough information to decide that what the couple has done is wrong – and that the world where the impaired child Andy exists is worse than the world where the nonidentical but better off child Ruth exists instead. The usual view is that the story also gives us enough information to conclude that what the agents have done in taking the pill is *maximizing for Andy*. And from there we are to infer that P* – in both its deontic and telic forms – is false.

In fact, however, that latter point – that the choice is *maximizing for Andy* – is not what the story tells us at all. The story conveys a good deal more detail than the modally impoverished graph itself suggests.

We need only think about the story in a little bit more detail to realize that it's clearly *possible* for Andy to have come into existence *without* the couple having taken the pleasure pill. Nor is the future where Andy exists and is better off simply a *logical* or *conceptual* possibility for agents but rather something agents had the power, ability and resources to bring about. It's, in other words, a possible future that was *accessible* to agents at the critical moment just prior to choice, a future that agents had the actual power, ability and resources to bring about, a future that is *not* barred, for example, by the laws of nature. After all, the pleasure pill *isn't* a *fertility* pill. The couple *could* have paused to take an aspirin instead – and then have proceeded toward conception on exactly the same schedule they in fact follow in *w1*.

Kavka agrees. The world where Andy exists and is better off is not completely out of reach. It's just, Kavka notes, enormously *improbable* that such a better-for-Andy world would have unfolded had the couple done otherwise than just as they did in *w1*.[31]

The view that probabilities bear on how choices are to be evaluated is, of course, entirely plausible. And we'll need to understand how Kavka and others

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

think that that point plays out in the pleasure pill case – why, that is, they think we are forced to conclude that what the couple has done doesn't make things worse for Andy – that is, that it is maximizing for Andy. But first let's correct the graph.

Type I nonidentity problem; modally enriched presentation			
Well-being	*w1* (including *a1*) what *does* happen (couple pauses to take pleasure pill)	*w2* (including *a2*) what *would have* happened (couple doesn't pause)	*w3* (including *a3*) what *could have* happened (couple pauses to take aspirin)
+10		**Ruth**	**Andy**
+8	**Andy**		
+0	*Ruth**	*Andy**	*Ruth**

One other preliminary point is in order before we turn to probability. It might seem that any discussion of probability will surely remain just a sideshow given that the case *stipulates* that, had the couple not taken the pleasure pill, Andy would never have existed at all. It might seem, that is, that, in view of that stipulation, *w3*, though *accessible*, must be *irrelevant* to the question whether *a1* makes things worse for Andy. It might seem that *a1* can make things worse for Andy – colloquially, *harm* Andy – *only if* had *a1* never been performed – only if *but for a1* – Andy would have been better off. And here by stipulation *but for a1* Andy would never have existed at all. Hence, whatever role probability might sometimes play in our analyses of other cases, we may seem forced in this case to conclude that Andy hasn't been made worse off. We may seem forced to say 'no harm done.'

But that way of understanding what it is for a given act to make things worse for, or to *harm*, a person has long been discredited. Suppose we are trying to figure out whether my shooting you in the arm makes things worse for you – whether it harms you. It's part of the case – I prove, we can imagine, in court – that, had I not shot you in the arm, I would have shot you in the head. Does that mean that I *didn't* harm you when I shot you in the arm – really, that I have *benefited* you, even perhaps that *you* owe something to *me* under a theory of unjust enrichment for my having so kindly saved your life? No, not, at least, on the further entirely natural and appropriate supposition that I also had the further option of *not shooting you at all*. That I *would* not have availed myself of that further option says something about *me* – about, perhaps, my character. But it *doesn't* show that I *haven't* harmed you when I shoot you in the arm. I have made you worse off, not compared against where you *would* have been had I not shot you in the arm, but against *another baseline altogether*: where you *could* have been and indeed *would* have been had I refrained from shooting you altogether.[32]

3.3. Argument from probability

The real obstacle, then, to any finding of 'harm done' in the pleasure pill case emerges not from a stipulation as to *what would have been* but rather from a

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

line of reasoning that exploits the surely correct fact that, for any person, that person's own coming into existence ordinarily remains, at least until the point of conception, a *highly improbable event*.

In the shoot-you-in-the-arm case, it's easy to see how I might have both *tried* to make and indeed *succeeded* in making things better for you. I could simply have stood there and not shot you at all. But what would have happened had the couple in the pleasure pill case *tried* to make things better for Andy? What would have happened had they diverted from their actual course in *w1* and *tried* to conceive the very same child they in fact conceive in *w1* by *trying* to pause to take the aspirin *on the very same schedule* they in fact follow in *w1*?[33] What would the *probability* have been that *w3* would have unfolded? What would the *probability* have been even that *a3*, a highly detailed *act* and not a generic *choice*, would have been performed in place of the many, many alternate acts that would have just as well implemented the couple's generic choice to pause to take the aspirin?

Almost nothing; close to zero. But surely it's better for Andy to have what he has in *w1* than it is for him to have only a *very tiny probability* of obtaining the better existence that he has in *w3*. Which surely in turn means that *a1* and *w1* are, after all, *maximizing* for Andy – and that P* is therefore false.

It's the inference to the result that *a1* and *w1* are *maximizing* that we should resist. We can't infer that *a1* and *w1* are maximizing for Andy on the basis of a comparison between the *actual* well-being Andy has in *w1*, given the couple's choice to pause to take the pleasure pill, against Andy's *expected* well-being in *w3* or in any other world, given the couple's choice to pause to take the aspirin instead. But *expected* well-being is what Andy's very low probability of ever existing at all, given the couple's choice to pause to take the aspirin in place of the pleasure pill, goes to. We might debate at length whether personal betterness is a matter of having more *actual*, or more *expected*, well-being. What we *can't* cogently do, however, is to mix and match actual well-being and expected well-being and then conclude anything about personal betterness at all.[34]

Philosophers who have considered my proposed strategy for dissolving the Type I problem have generally resisted it, pointing out that we'd surely all prefer to retain the flawed but good existence we in fact have rather than relinquish what we have in exchange for a very small chance at a still better existence.[35]

If there were *any point* in the case where the temporal ordering of Andy's alternatives lined up in just that way, then that point might gain some traction. But there isn't. The cases are structurally distinct.

Consider, e.g., the moment just prior to choice. The couple's choice to pause to take the pleasure pill at *w1* doesn't *secure* Andy's existence. Just prior to choice, it's no more probable that Andy will exist given the couple's choice to pause to take the *pleasure pill* than that Andy will exist given the couple's choice to pause to take the *aspirin*.[36] I concede that, just prior to choice, the probability that Andy will come into existence given the couple's choice to pause to take

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

the aspirin is very small. But clearly, just prior to choice, the probability that Andy will come into existence given the couple's choice to pause to take the pleasure pill is also very small and indeed just as small. But the *outcome* for Andy, if he does exist, is considerably better under the one choice than under the other. A standard expected value calculation then tells us that the choice to pause to take the pleasure pill doesn't, after all, maximize expected well-being on behalf of Andy.

The upshot? The couple's taking the pleasure pill maximizes neither actual nor expected well-being on Andy's behalf. Thus, we can, whether we think that making things worse for Andy is a matter of diminished *actual* well-being, diminished *expected* well-being or *both*, find ample room to conclude that their taking the pleasure pill does, after all, make things worse for Andy.[37]

I believe, accordingly, that we have dissolved the Type I problem. We have a clearly wrong act and a clearly worse world. But that act and that world are also *worse for* Andy. We are thus left with no grounds for rejecting P*.

4. Objections to P*

4.1. Why P* is critical to proposed solution to the nonidentity problem

I noted earlier that virtually all theorists formulate the core person-affecting intuition in a way that allows us to reach the critical inference after a significantly narrower inquiry than the more far-reaching inquiry that P* demands.[38] We can put the telic component of the standard formulation – which I'll call Ps (telic) – as follows.

> Ps (telic): A world *w* that is worse than another world *w*′ must be worse for a person who does or will exist at *w* than *w*′ is.

Thus, P* requires us, before we can determine that *w* isn't worse than an alternate world *w*′ is, to canvas *all* accessible alternates *w*″ to *w*. If *w* is worse for any person *p* who does or will exist in *w* than *any such w*″, then the necessary condition on outcome worseness that P* sets forth is satisfied and we can't infer from P* that *w* isn't worse than *w*′ is. In contrast, Ps generates the result that *w* isn't worse than *w*′ is far more readily – whenever there is no *p* in *w* such that *w* is worse for *p than w*′ *is*.

The distinction here is not merely technical. P*'s more far-reaching inquiry is essential to the person-affecting solution to the Type I problem that I have proposed.

Thus, about the pleasure pill case, I want to say that *w3* shows that *w1* is worse not just than *w3* is but also worse than *w2*. We need to say *w1* is worse than *w3* to solve the nonidentity problem. But we also need to say that *w1* is worse than *w2*. Why? It's highly implausible that *w2* is either better or worse than *w3* is. (Notably, P* concurs.) But if that's so, then, since *w1*, we agree, is worse than *w3*, we must also agree that *w1* is worse than *w2*.[39] But Ps itself says that it's not;

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

Ps implies that it's not the case that *w1* is worse than *w2* since there's no one in *w1* such that *w1* is worse for that person than *w2* is. In contrast, P* opens the door to our saying exactly what we need to say about that case: that *w1* is worse than both *w2* and *w3* and that *w2* is exactly as good as *w3* is.

4.2. Objection based on axiology; inconsistency argument

But does the more far-reaching P* run afoul of a certain axiological constraint that Ps itself survives? The clearest argument that it does comes in the form of an *inconsistency* argument.

Consider the following two-outcome case one in which only *w2* is accessible to *w1*:

Two-outcome case		
Well-being	*w1* (including *a1*)	*w2* (including *a2*)
+10		
+8	**p**	
+0		*p**

P* (telic) implies that *w1* isn't worse than *w2* since there's no existing or future person in *w1* such that that person is worse off in *w1* than that person is in any alternate accessible outcome. But now consider a three-outcome case that leaves everything just as it is in the two-outcome case but adds a third accessible outcome.

Three-outcome case			
Well-being	*w1* (including *a1*)	*w2* (including *a2*)	*w3* (including *a3*)
+10			**p**
+8	**p**		
+0		*p**	

P* implies that *w2* is at least as good as *w3* is – indeed, that *w2* and *w3* are equally good. Those results are perfectly plausible. But as noted earlier we want and need to say as well that *w1* is worse than *w2* is. Since there is someone, *p*, who is worse off in *w1* than that same person is in *some alternate accessible outcome*, namely, *w3*, the condition P* (telic) spells out on *w1* being worse than *w2* is satisfied. P* thus allows us to avoid the result that *w1* is at least as good as *w2* is, which in turn leaves us room to say that *w1* is worse than *w2*.

Of course, we draw upon a further person-affecting principle to generate that latter result; we can't on the basis of just P* conclude that *w1 is* worse than *w2*. But that further principle will, like P*, exploit the fact that *w1* is worse for some person *p* who does or will exist in *w1* than *w3* is for *p*.[40]

But now we have an inconsistency. We have said both that *w2* is at least as good as *w1* – that's the result we obtained when we applied P* (telic) to the two-outcome case. But applying both P* along with our further principle to the three-outcome case, we then obtained the result that *w1 is* worse than *w2*.

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

Ps safeguards against such an inconsistency. It makes *w3* irrelevant to our comparison between *w1* and *w2*. It insures that what we say about how *w1* compares against *w2* in the two-outcome case *just is* what we say about how *w1* compares against *w2* in the three-outcome case. It insures that what we say in both cases is that *w1* is at least as good as *w2* is.

Now that seems a problematic result – and sufficient grounds, I think, on which to reject Ps. But Ps is not my concern here. My concern here is to determine whether P* indeed involves us in inconsistency – a result that would render what I consider a highly plausible solution to the Type I nonidentity problem impossible.

Is there then a way to retain P* but avoid the inconsistency? In the foregoing, I have been explicit that what we are comparing are not mere *distributions* of well-being across identified populations but rather full-blown, fully detailed, possible *worlds*, or *futures*, or *outcomes*. So let's think about those details for a moment. In the two-outcome case, agents lack the power, the ability, the resources, to make things better for *p* than they are in *w1*. Their *only options* in *w1* are to bring *p* into existence at the lower well-being level and to leave *p* out of existence altogether. In contrast, in the three-outcome case, agents have exactly that power, that ability, those resources, in *w1*. The agents there are fully capable of avoiding the loss *p* incurs in *w1* on *p's* behalf in the three-outcome case.

But worlds are finely distinguished; any little 'change' in a given world means we are talking about a different world altogether. If I have an ability, a power, a resource in one world that I lack in another, then we can be sure that the worlds referred to constitute *two distinct worlds*.[41]

We are thus free to say, indeed it seems *must* say, that *w1* and *w2* in the two-outcome case just aren't the same worlds as *w1* and *w2* in the three-outcome case. We avoid the inconsistency by introducing a more exacting vocabulary. In the three-outcome case, the fact of the third outcome is in effect entailed by our descriptions of the first two worlds. In the two-outcome case, our descriptions of the first two worlds rule out that third world.

The upshot? It remains a *complete mystery* to me why theorists associate Ps with the person-affecting view. If Ps is part of the person-affecting view, then the person-affecting view can't solve the nonidentity problem. But why not go with P* instead?[42]

4.3. Objection based on accessibility

A further concern about P* arises from the fact that P* – along with the family of person-affecting principles that P* seems naturally to fit in with – ties its worseness rankings to a notion of when one world is *accessible* to another. I have here adopted a fairly liberal notion of *accessibility*. For purposes here, *accessibility* has

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

been defined to include, not just those outcomes the *individual* agent, working on his or her own and independently of what other agents do, has the power, the resources and the ability to bring about, but also those outcomes the agent *working together with other agents* (whether in concert or collaboration or by agreement or simply coincidentally) has the power, the resources and the ability to bring about.[43]

The problem with building the narrower notion into our formulation of the telic form of the person-affecting intuition is that the resulting principle generates multiple worseness rankings – rankings that we would need to relativize in order to avoid inconsistency – depending on whether we (a) focus on just the outcomes that the *individual* agent has the power, the resources and the ability to bring about, or (b) focus on outcomes that agents *working together* have the power, the resources and the ability to bring about.[44] And it seems plausible that relativizing our rankings in that way would be a mistake.

The worry is that our more liberal notion of *accessibility* will mean that the connection between the permissibility of a single individual agent's act and outcome worseness – between the *deontic* and the *telic* projects – may not be as clear-cut as we may have thought (or hoped) it was. On the other hand, the defense '*Whatever* act I performed, given the acts of other agents, the same *horrible* outcome would have obtained; therefore, *my* act was *permissible*' often fails in spectacular ways. Participating in a firing squad consisting of nineteen other armed agents isn't necessarily permissible, even if *not* participating would not have produced a better outcome.

So perhaps it's not, in the end, clear that a more liberal notion of *accessibility* is problematic. Perhaps indeed an approach that casts a wider net on outcomes that are *accessible* – and then carves out from the class of otherwise impermissible acts those acts performed by an agent as an individual in cases where what that agent has done by participating in bringing about the bad outcome actually produces a *better* outcome than the outcome produced by that agent's simply walking away – will in the end seem a more workable approach.[45]

5. Addition plus

As we have seen, understanding the person-affecting approach to include P* and not Ps (and certainly not Px) is critical to providing a person-affecting solution to the nonidentity problem. It's worth noting that understanding the person-affecting approach in that way may also represent our only hope for providing a person-affecting solution to addition plus – or indeed to the mere addition paradox itself. My suggestion, more generally, is that a successful person-affecting solution to any such problems requires recognition that how two outcomes compare may be determined in part by what is going on at a third outcome.

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

Addition plus			
Well-being	*w1* (including *a1*)	*w2* (including *a2*)	*w3* (including *a3*)
+10		**p**	
+9	**p**		
+5			**p, q**
+1		**q**	
+0	*q**		

It may be that theorists don't turn to P* for purposes of analyzing cases like addition plus because they are concerned, *sotto voce*, about the very inconsistency argument discussed in Section 4.2 above.

Consider, for example, how de Lazari-Radek and Singer analyze addition plus. At one point in their argument, they introduce just the two worlds – which I label *w1* and *w2* here – and ask us to compare those worlds. It's made clear that we aren't, at that point, to imagine *w3* lurking in the background as an available option for agents. Thus, in comparing *w1* against *w2*, we're to suppose 'natural phenomena [that] cannot be changed'; we're to suppose that 'compensation [for *q*] is impossible.' On those facts, I am happy to say – with de Lazari-Radek and Singer and as directed by P* – that *w2* is at least as good as *w1*. Thus, Lazari-Radek and Singer write that they can see 'no good reason for preferring [*w1*] to [*w2*].'[46]

A few pages later, de Lazari-Radek and Singer 'add a third possibility' to the case, *w3*.[47] We are then asked to compare *w3* against *w2* and conclude that *w3* is at least as good as *w2* is – and indeed that it's better. I have no quarrel with those results either – and nor does P*.

They then, however, proceed to do something that *does* create a problem. They import without further discussion the result that we agreed to at the beginning, where we were just asked to compare *w1* against *w2* and were to suppose that 'compensation' to *q* in *w2* was 'impossible' and that 'natural phenomena' could not be changed: that *w2* is at least as good as *w1* (this is just a reminder what we said before). Transitivity then, as they note, produces the result that *w3* must be at least as good as *w1* is. But that's a result they – and we all, I think – want to reject, not in the least part because not to reject that result seems to take us in the direction of the repugnant conclusion. We want to say instead that *w3* is worse than *w1*. Can we do that?

Not in a way that is consistent with the line of reasoning we have just worked through. de Lazari-Radek and Singer themselves do not suggest a way out. They, rather, simply forewarn us of the repugnant conclusion.

But I think we do have a way out. I believe that when de Lazari-Radek and Singer 'add a third possibility' – that is, the third possible outcome or future or world *w3* – to the case, they in fact substitute in, for the original pair of worlds they have us consider at the beginning of their discussion, a *numerically distinct* pair of worlds. Where 'compensation' for *q* wasn't 'possible' before, now it is. Where 'natural phenomenon' before could not be 'changed' to make things better for *q*, now they can. New facts mean, not that there was anything amiss

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

in our *original* comparison between the two worlds that we were *originally* presented with, but that we need to realize that those original worlds are *not part of the new case*. About the two worlds we are *now* being asked to compare, then, we can consistently take the position that – given the third world – the second world *is* worse than the first.

We simply recognize, in other words, that the worlds we were asked to compare starting out – where 'compensation' to *q* was 'impossible' and 'natural phenomena' can't be changed – just *aren't the same worlds* as the worlds we are talking about at the end. A *more exacting vocabulary* allows us to avoid both confusion and inconsistency. It allows us to recognize that the wide-angled P* itself may consistently treat the two cases differently, which gives us the room we need, in turn, to say that for addition plus – a case that *includes* the third outcome; a case in which compensation to *q is* possible and natural phenomena *can* be changed – the accessibility of *w3* makes *w2* not just worse than *w3* but worse than *w1* as well.

Temkin, famously, has found his way out of the mere addition paradox by denying transitivity.[48] If, however, we recognize that how two outcomes compare may be determined in part by facts about other accessible outcomes, we can escape the inconsistency without denying transitivity.[49]

6. Conclusion

Recognizing that how two outcomes compare may be determined in part by facts about *still other accessible worlds* may free our hands to provide a plausible person-affecting solution to the nonidentity problem, to addition plus and perhaps to other problems in ethics as well. I have here focused on certain axiological considerations that may seem to rule out any such wide-angled approach to determining when one world is worse than another. And I have argued that the inconsistency that that approach may seem quickly to involve us in can be resolved by appeal to a more exacting vocabulary, one that recognizes such other accessible worlds as critical to determining, not just how one world compares against another, but the very identities of the worlds themselves. And I have also argued that the liberal notion of accessibility it seems the more wide-angled approach requires is not itself obviously problematic and may indeed help us analyze problems beyond the nonidentity problem itself.[50]

Notes

1. I have elsewhere proposed substantially the same solution to this type of nonidentity problem. See Roberts (2007, 2009). It has not been widely endorsed, perhaps as a result of the objections I consider, and argue fail, in Section 4 below.
2. Temkin (1987, 2012).
3. 'Utilitarianism does not take seriously the distinction between persons.' Rawls (1972, 27).

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

4. For discussion of the separateness objection to totalism, see de Lazari-Radek and Singer (2014, 137–140).
5. Feldman thus introduced a concept of *adjusted utility*, which he proposed as helpful in solving the repugnant conclusion without rejecting aggregation. See Feldman (1995a and 1995b). Broome's concept of the *personal good* functions similarly. Broome (2015).
6. This is not to say that a person's *moral* status depends on his or her *modal* status. Rather, it's to note that the moral status of a given *loss* might depend on whether the person who incurs the loss does or will exist at the world where the loss is incurred. See Section 2.5 below; see also Roberts (2011a, 2011b).
7. To insist that we shouldn't aggregate *well-being* across the population is not to accept that we shouldn't *aggregate*. It's just to accept that, if we aggregate at all, we should aggregate *something other than* that which makes the particular existence at the particular world so precious to – or so miserable for – the one who exists.

   Thus, we can take the position that what is to be aggregated across the population at a given world is not *well-being* but rather what Broome calls the *personal good*, which, according to Broome, we are free to define in a way that recognizes the values of equality, priority and fairness. Broome (2015). We might, for example, take the position that a low well-being level that one suffers as a result of an inequality in the distribution of well-being across the population yields a personal good that is lower still (lower, that is, than we might have thought it would be given the well-being level alone).

   I have argued elsewhere that, if the concept of the personal good is indeed *that* flexible, then it is also flexible enough to accommodate our *existential* values as well, that is, to recognize that the positive well-being level one has at a world where one exists and is happy may itself translate, not to a *positive* level of the personal good, but rather to a *zero* level. On that view, the existence of the additional happy person doesn't make the world morally better. See Roberts, *Modal Ethics*, Part II (unpublished manuscript).

   That latter thesis is sometimes associated with what Broome calls the *neutrality intuition*. Broome (2004). According to the neutrality intuition, within a certain 'neutral range,' a given existence doesn't make the world either morally better or morally worse. Broome argues, however, that the neutrality intuition leads to inconsistency. I accept Broome's argument but point out that we might reject the neutrality intuition itself but accept an alternate principle, that is *narrow neutrality*. According to narrow neutrality, the existence worth having doesn't make the world morally better but does, often, make it worse. For additional discussion, see Roberts, *Modal Ethics*, Part II.
8. Thus, for purposes here, I accept that the *deontic* and the *telic* projects are linked in important ways. In general, we determine the permissibility of a given act performed at a given world by reference to the other possible worlds (or futures or outcomes) that agents have the resources, ability and power to bring about in place of the one; the accessibility of any *better* alternate outcome means a *wrong* act. I also think, however, that the fact that an agent, *working on his or her own*, does not have the resources, the ability or the power to bring about a better outcome does not necessarily mean that what the agent as an *individual* has done is permissible. Participating in a firing squad consisting of nineteen other armed agents isn't necessarily permissible, even if *not* participating would have failed to bring about any better outcome. For further discussion, see Section 4.3 below.

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

I take for granted here that the term *person* includes many non-human animals but may well exclude some humans.

9. This is not to say that we won't in the end aggregate. See note 7 above. It is just to say that we'll first determine what well-being gains and well-being losses are relevant to whether one world is morally worse than another. And only then might we aggregate – not *well-being*, but rather what Broome calls the *personal good*.
10. Now, the fact that our principle explicitly includes that limitation may create some confusion. If our *core* principle explicitly limits the range of people the condition applies to, then it might be assumed that *all our further* person-affecting principles must do so as well. But that's just an assumption. And – as we shall see in Section 2.4 below – it's an assumption we should resist.
11. For further discussion of *accessibility*, see Sections 3.2 and 4.3 below; see also Feldman (1986, 10–11 and 16–25). The notion of *accessibility* I use here owes much to but is also more liberal than the concept Feldman has developed. For my purposes, if agents working together at a world *w* have the power, resources and ability to bring about an outcome *w'*, then *w'* is *accessible* to *w* even if an *individual* agent had no means of making things any better than they in fact turn out to be.
12. Parfit ([1984] 1987, 363).
13. See e.g. Temkin (1987, 167), cl. (ii) of what he calls the 'person-affecting principle' (or 'PAP'); Temkin notes that he bases his own formulation on Parfit's.
14. For an example of how the telic component of the person-affecting intuition is standardly formulated, see Holtug (2010, 158).

    Holtug and many other philosophers go even further: *w* is *better* than *w'* only if there is a person who does or will exist in *w* and *w* is *better for* that person than *w'*. Holtug (2010, 158), 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 *p* who does or will exist in the one outcome and the one outcome is better for *p* than the other outcome is itself subject to instant counterexample. Thus, in cases involving the miserable child half of the procreative asymmetry, 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 is 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.

    But even eliminating the references to *betterness* and restricting the principle so that it functions just as a necessary condition on when *w* is *worse than w'* fails to save the principle. It won't, then, be subject to instant counterexample. But it will still fail – for my purposes, at least – by virtue of the fact that it forces the result that one world isn't worse than another on the basis of an unduly narrow inquiry – and, as such, would completely undermine any attempt to forge a plausible person-affecting solution to the nonidentity problem. See Section 4.1 below; see also Roberts (2015).

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

It's worth nothing that philosophers who focus on the deontic component of the person-affecting intuition also often (if perhaps not quite standardly) formulate that intuition as well in a way that deems an act permissible on the basis of an unduly narrow inquiry. Thus: *a1* performed at *w1* is wrong, *only if* there does or will exist a person *p* in *w1* such that had *a1* not been performed, *p* would have been better off. See e.g. Dasgupta (forthcoming), part 1 (combination of his claims (2) and (3)). Equivalently, the deontic intuition may be spelled out in terms of *harm*, where *harm* itself is then defined in a defective way (most commonly, by reference to a simple counterfactual 'but for' account of *harm*). See Boonin (2015, 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, 8). The shoot-you-in-the-arm case shows, I believe, that such accounts are problematic (whether or not they are formulated in terms of *harm*). See Section 3.2 above.

15. Thus, a workable Pareto-like person-affecting sufficient condition on when one world is worse than another includes the following:

    P** (telic): A world *w* is worse than another world *w'* if there is a person who does or will exist at *w* such that *w* is worse for that person than *w'* is and there is no person who does or will exist in *w'* such that *w'* is worse for that person than any accessible world *w''* is, *where w'' may but need not be identical to w*.

    This principle can be contrasted with still another Pareto-like principle, Px. I describe Px in Section 2.4 below but immediately reject it.

16. Temkin is probably best understood as intending to include Px as clause (i) in his own formulation of the 'person-affecting principle' (or 'PAP'). See Temkin (1987, 166).
17. de Lazari-Radek and Singer (2014, 367).
18. It might be argued that Px yields that result only on the assumption that it is sometimes better for a particular person never to have existed at all. I do not think, however, that that assumption is one we should contest.
19. de Lazari-Radek and Singer (2014, 367–368).
20. de Lazari-Radek and Singer (2014, 368).
21. Roberts (2011a, 2011b).
22. If might be objected that, if the nexus between existence and *loss* makes things *worse*, then the nexus between existence and *gain must* make things *better*. It seems, however, that to defend variabilism against this objection, we need simply deny that the implication *must* in fact hold. Thus, variabilism asserts that the nexus between existence and loss and only that nexus makes the person's loss morally significant. It follows – as a conceptual matter, given that a *loss* a person incurs at a world *w* compared against another world *w'* is identical to the *gain* that same person accrues at *w'* compared against *w* – that certain gains are also morally significant – but only those that *reverse* morally significant losses. There's nothing self-contradictory in that view – unless we import into that view *some further view*, one that insists that the nexus between existence and gain makes the person's gain morally significant. When we do that, we end up with a view that contradicts itself. But *that's* not the view that, starting out, was under attack or that we aimed to defend.

    Variabilism may legitimately be considered a version of the *complaint view*. But it's a non-Singer-esque version of that view. Thus, the objection de Lazari-

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

Radek and Singer put forward against the (Singer-esque) version of the complaint view that they explore *doesn't* work against variabilism. (See de Lazari-Radek and Singer (2014, 269–370).) Variabilism isn't, that is, applied to addition plus, committed to the position that, in virtue of the applause agents get in *w3*, that *w3* is better than *w1*. Variabilism, rather, is committed only to the position that the applause agents get in *w3* shows that *w3* is better than *w2*. We should agree that *w3* is better than *w*2 in virtue of the fact that *w*3 is better in a morally significant way for *q*. But we can still deny that *w3* is better than *w1* by denying that *w3* is better for *q* than *w1* is in a way that has *any moral significance at all*.

23. See Roberts (2007 and 2009); see also Roberts and Wasserman (2017).
24. I have called it the *nonidentity fallacy*. See note 23 above.
25. Some philosophers favor a non-comparative account of *harm* and argue, accordingly, that harm, or loss, can be imposed in a case where the person's well-being has been maximized – even, that is, in a Type II case. But while I am happy to say that the term 'harm' may well have such a non-comparative sense – along with many others; 'harm' seems to me to be a multiply ambiguous term – I find it implausible that *harm* in that sense has any moral significance. It seems plausible to me, in other words, that, for a harm or a loss or a simple diminution in well-being to have *moral significance*, the option of things being at least somewhat *better* for the person must exist.
26. My thinking on this point has been greatly influenced by David Wasserman. See note 23 above.
27. Kavka (1981, 98). Explaining the 'precariousness' of existence, Kavka, citing Adams, Parfit and Schwartz, writes: 'Which particular future people will exist is highly dependent upon the conditions under which we and our descendants procreate, with the slightest difference in the conditions of conception being sufficient, in a particular case, to insure the creation of a different future person … This fact forms the basis of a surprising argument …' Kavka (1981, 93).
28. Kavka (1981, 98).
29. *Andy* thus is a perfectly good genuine proper name; *Ruth* a constant.
30. Now, Kavka doesn't himself rely on that particular counterfactual. Other theorists, however, do; and we can accept it here as a stipulation of the case: in the closest possible world where the couple doesn't take the pleasure pill, Andy never exists.
31. See Kavka (1981, 100, n. 15).
32. This is just to say that the simple counterfactual 'but for' test of *harm* – that is, of making things *worse for* – is false. We must reject the principle that asserts that *a1 harms* a person *p* only if had *a1* not been performed – *but for a1 – p* would have been better off. The concept of the baseline here is Feinberg's. Feinberg (1988, 150–151). I question, however, the simple counterfactual account of harm he explores.
33. Of course, if we want to further complicate their efforts, we can simply note that they would have had no way of knowing just which child that would have been.
34. And it's a fallacy to think that we can. It's not implausible that what people have been doing with this case and other cases that ground the Type I problem is comparing the subject's *actual* well-being in *w1* given *a1* against the subject's *expected* well-being under various alternate choices. Now we might think the important thing is whether a *choice*, *expectationally*, is worse than another for a person, in which case we should compare expected value compared against expected value. Or we might think the important thing is whether a *world* is *in fact* worse than another for a person, in which case we consider actual value

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

compared against actual value. But we can't mix things up without facing inconsistency. See note 23 above.

35. Smilansky (2017).
36. It's the *pause*, not the *pill*, that arguably gives Andy whatever chance he has of coming into existence. And, whatever the duration of that pause turns out in fact to be at *w1*, the probability, just prior to choice, that that pause will have *exactly that duration* remains very small.
37. We have ample room to reach those felicitous results, in other words, on the basis of, for example, the person-affecting principle P** (not Px). See note 15 above.
38. See note 14 above (discussion of Fleurbaey etc.).
39. I here assume a tripartite theory of outcome betterness: if *w2* is neither better nor worse than *w3*, then *w2* is equally as good as *w3*.
40. That same principle – P**, not, of course, Px – is the same principle we need to solve the nonidentity problem. See note 15 above. In contrast, Ps would close the door to that further principle by forcing us to say that *w1* is at least as good as *w2* is.
41. In taking this approach, we aren't eschewing pairwise comparison. It's just that when we conduct our pairwise comparisons we have to take facts about accessibility into account. We look at *w1* – and see that for agents there *w3* remains a perfectly live option.
42. The discussion of this Section 4.2 shows that the wide-angle-lens approach to determining betterness we see in P* doesn't, after all, violate the independence principle. For this approach simply denies that there could be both a case in which *w3* exists as an accessible alternative from *w1* and *w2* and a case in which *w3* doesn't exist as such an alternative; one of the two cases would be ruled out as impossible in view of our understanding that each world has its features *necessarily*. Specifically, the following accessibility axiom seems both undeniable, given our understanding of worlds, and plausible.

    Accessibility axiom: If $w\beta$ is accessible to $w\alpha$, then *necessarily* $w\beta$ is accessible to $w\alpha$.

43. See note 11 and Section 3.2 above (on accessibility).
44. Now, in the context of the nonidentity problem, the individual agent often can, on his or her own, bring about the better outcome. Thus, in the pleasure pill case, whoever it is who ingests the pleasure pill was fully capable of ingesting the aspirin instead; the acts of other agents do not block that choice. But we can certainly sketch nonidentity cases where that condition is not satisfied – problems, in other words, that *combine* the nonidentity problem *and* a collective action problem. In any case, the goal here is not simply to address the nonidentity problem but also to develop principles that stand up to scrutiny in other sorts of cases as well.
45. I am grateful to Elizabeth Harman for bringing to my attention to the need to address this issue here.
46. de Lazari-Radek and Singer (2014, 366).
47. de Lazari-Radek and Singer (2014, 370).
48. See Temkin (1987, 2012).
49. McMahan has worried about similar inconsistencies, proposing that our betterness results will themselves depend on our 'order of comparison.' See McMahan (2006). On the approach suggested here, we shall have no need to relativize our results in that way. If a third world *w3* – as in addition plus or the mere addition paradox – shows that *w2* is itself morally defective, that defect

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

shall stay with *w2*, whether it is *w3* we are comparing *w2* against or *w1* we are comparing *w2* against. The order in which we consider things is not important; what is important is that we not think we can accurately compare *w2* against *w1* without first understanding whether *w3* exists as an accessible alternative.

50. I am very grateful to Rahul Kumar for his comments on an earlier version of this paper. I presented a still earlier version of the paper at a Population Ethics Workshop organized by the Institute for Futures Studies (Stockholm, May 2016), and I am very grateful for the comments of those who participated in that conference as well.

References

Arrhenius, Gustaf. 2015. "The Affirmative Answer to the Existential Question and the Person Affecting Restriction." In *Weighing and Reasoning: Themes from the Philosophy of John Broome*, edited by Iwao Hirose and Andrew Reisner, 110–125. Oxford: Oxford University Press.

Boonin, David. 2015. *The Non-Identity Problem and the Ethics of Future People*. Oxford: Oxford University Press.

Broome, John. 2004. *Weighing Lives*. Oxford: Oxford University Press.

Broome, John. 2015. "General and Personal Good: Harsanyi's Contribution to the Theory of Value." In *The Oxford Handbook of Value Theory*, edited by Iwao Hirose and Jonas Olson, 249–266. Oxford: Oxford University Press.

Dasgupta, Shamik. Forthcoming. "Essentialism and the Nonidentity Problem." *Philosophy and Phenomenological Research*.

de Lazari-Radek, Katarzyna, and Peter Singer. 2014. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press.

Feinberg, Joel. 1988. "Wrongful Life and the Counterfactual Element in Harming." *Social and Political Philosophy* 4 (1): 145–178.

Feldman, Fred. 1986. *Doing the Best We Can*. Dordrecht: Reidel.

Feldman, Fred. 1995a. "Adjusting Utility for Justice: A Consequentialist Reply to the Objection from Justice." *Philosophy and Phenomenological Research* 55: 567–585.

Feldman, Fred. 1995b. "Justice, Desert, and the Repugnant Conclusion." *Utilitas* 7: 189–206.

Fleurbaey, Marc, and Alex Voorhoeve. 2015. "On the Social and Personal Value of Existence." In *Weighing and Reasoning: Themes from the Philosophy of John Broome*, edietd by Iwao Hirose and Andrew Reisner, 95–109. Oxford: Oxford University Press.

Holtug, Nils. 2010. *Persons, Interests and Justice*. Oxford: Oxford University Press.

Kavka, Gregory. 1981. "The Paradox of Future Individuals." *Philosophy and Public Affairs* 11: 93–112.

McMahan, Jeff. 2006. "Paradoxes of Abortion and Prenatal Injury." *Ethics* 116 (4): 625–655.

Mulgan, Tim. 2006. *Future People: A Moderate Consequentialist Account of our Obligations to Future Generations*. Oxford: Clarendon Press.

Parfit, Derek. [1984] 1987. *Reasons and Persons*. Oxford: Oxford University Press.

Rawls, John. 1972. *A Theory of Justice*. Oxford: Oxford University Press.

Roberts, Melinda A. 2007. "The Nonidentity Fallacy: Harm, Probability and Another Look at Parfit's Depletion Example." *Utilitas* 19: 267–311.

Roberts, Melinda A. 2009. "The Nonidentity Problem and the Two Envelope Problem." In *Harming Future Persons*, edited by Melinda A. Roberts and D. Wasserman, 201–228. Dordrecht: Springer.

Roberts, Melinda A. 2011a. "The Asymmetry: A Solution." *Theoria* 77: 333–367.

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press

Roberts, Melinda A. 2011b. "An Asymmetry in the Ethics of Procreation." *Philosophy Compass* 6 (11): 765–776.
Roberts, Melinda A. 2015. "Population Ethics." In *The Oxford Handbook of Value Theory*, edited by Iwao Hirose and Jonas Olson, 399–423. Oxford: Oxford University Press.
Roberts, Melinda, and David Wasserman. 2017. "Dividing and Conquering the Nonidentity Problem." In *Current Controversies in Bioethics*, edited by Matthew Liao and Collin O'Neil, 81–98. New York, NY: Routledge.
Smilansky, Saul. 2017. "The Nonidentity Problem: United and Strong." In *Current Controversies in Bioethics*, edited by Matthew Liao and Collin O'Neil, 99–114. Routledge.
Temkin, Larry. 1987. "Intransitivity and the Mere Addition Paradox." *Philosophy and Public Affairs* 16 (2): 138–187.
Temkin, Larry. 2012. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford: Oxford University Press.

https://doi.org/10.1080/00455091.2017.1279518 Published online by Cambridge University Press