

## Chapter 2

# The Moral Significance of Merely Possible Persons

### 2.1 Who Matters Morally?

2.1.1 It is a matter of dispute whether persons who count as *existing or future*, relative to a particular circumstance, or possible future or *world*, have a moral status that *merely possible persons* lack. Are persons who do or will exist the only persons who matter morally? Are they the only persons whose needs and interests we must take into account in calculating what we ought to do? Can only their losses *bear on – count against* – the permissibility of acts that impose those losses? Can only their losses, in a roundabout way, *count in favor* of alternative acts that avoid those losses?

Or is the better view that *merely possible persons* – persons who will never exist at all – matter morally too? Must we take their needs and interests into account, alongside our own, in determining what we ought to do? Do losses incurred by merely possible people – like the losses incurred by you and me – *bear on* the permissibility of the acts that impose those losses? Can the losses they incur make the otherwise permissible acts that impose those losses *wrong*? Can avoiding a loss on behalf of the merely possible make otherwise wrong acts *right*?

My first goal in this Chapter 2 is to show that the radically exclusive position regarding the moral significance of merely possible persons – what I will call *Exclusion* – and the radically inclusive position – *Inclusion* – are both completely untenable. My second goal is show that there exists middle ground between these two extreme positions, and that this middle ground, what I will call *Variabilism*, is the far more plausible position.

I then briefly discuss what John Broome calls the *Neutrality Intuition* and what Peter Singer calls the *Prior Existence View*. The *Asymmetry* is a good, final test case for Variabilism, and I conclude this chapter with a note on that puzzle.

2.1.2 The view that we are to *exclude* how merely possible persons are affected by a given act and its alternatives in our calculations of what we ought to do – the view, that is, that the merely possible do not matter morally – is sometimes called

*Moral Actualism*.<sup>30</sup> Some theorists might be attracted to *Moral Actualism* because they are attracted to *Modal Actualism*. The thought is that, if there are no nonactual persons, then there is no need for moral law to concern itself with their plights. But it seems that we can reject the metaphysical doctrine while retaining the moral doctrine and vice versa. Moreover, *Moral Actualism* itself comes in two forms, Strong and Weak.<sup>31</sup> But only one of those forms is strictly *actualist*. Only one form takes the position that persons matter morally *if and only if* they do or will exist at the uniquely *actual* world. The other has us say instead that persons matter morally *if and only if* they would have existed – that is, *would have been* “*actual*” – had the act under scrutiny been performed.<sup>32</sup>

For those reasons, I will leave the term *Moral Actualism* (Strong and Weak) behind and use simply *Exclusion* (Alpha and Beta) instead.

2.1.3 Independent, then, of our metaphysics, *Exclusion* may immediately strike us as commonsensical. As between *Inclusion* and *Exclusion*, at least, it is *Exclusion* that seems to have the capacity to recognize the critical moral distinction between “making people happy” and “making happy people.”<sup>33</sup>

*Inclusion*, by comparison, seems utterly fantastic. According to *Inclusion*, we must *include* how the merely possible are affected, right alongside how we ourselves are affected, in making our calculations of what we ought to do.<sup>34</sup> The problem is that – as we

---

<sup>30</sup> See Josh Parsons, who calls the view *Axiological Actualism*, and Rivka Weinberg (unpublished manuscript). Parsons (2002), pp. 135–147. Jan Narveson, too, is sometimes considered an advocate of *Moral Actualism*. See Narveson (1976). See also David Heyd (2009).

Caspar Hare suggests that *Moral Actualism* – *Exclusion* – is tantamount to the person-based approach, which includes (among other things) the person-based intuition. See Hare (2007). In fact, however, *Exclusion* is just one way of articulating a person-based approach. Variabilism is an alternate and far more tenable way of articulating both the intuition itself and the approach. See note 17 above and part 2.2.4 below.

<sup>31</sup> Hare uses these terms. See Hare (2007), pp. 502–503.

<sup>32</sup> Many theorists find modal actualism an attractive view. Very stringent forms of that view may seem to compel us to adopt *Moral Actualism* (Alpha) or to dismiss as nonsense any effort to say that a world at which a person exists can be better (or worse) for that person than a world at which that person never exists at all. But such very stringent approaches seem problematic. A plausible semantics, actualist or not, must make sense of the sentence “JFK could have had another child who was a senator but could have been an astronaut instead.” See McMichael (1983). See also Rosen (1990) and Sider (2002).

<sup>33</sup> Narveson (1976), p. 73.

<sup>34</sup> Again, many might object to this way of talking. Surely the merely possible do not in fact *have* interests in the same way you and I might have interests in, for example, having the thermostat turned down. To avoid begging the question in favor *Inclusion* – or against *Inclusion*, by making *Inclusion* sound like the silly idea that we must knit and bake chocolate pies for a person who will never exist at all – it is critical not to read any particular substantive moral position into this way of talking about the merely possible. We can, in other words, talk about the merely possible as “having interests” or “incurring losses” – even if at the end of the day we conclude, with the *Exclusionists*, that those interests and losses are completely devoid of moral significance or, with the *Variabilists*, that some of those interests and losses have moral significance but some do not. In short, the purpose of talking this way is to achieve additional clarity, not to beg the question.

shall see – such “voices from another world” are hardly mellifluous.<sup>35</sup> “Bring me into existence,” they wail – even when for that to be done *real, live, flesh-and-blood* persons must be made to suffer. But how can my choice to make things better for you be wrong when all I have done to make things better for you is leave someone else out of existence altogether?

2.1.4 Later in this chapter, I detail just how extreme Inclusion in fact is. But as it happens Exclusion is just as extreme. If Inclusion takes the merely possible too seriously, Exclusion – as we shall also see – does not take them nearly seriously enough. The upshot is that Exclusion, when combined with an otherwise plausible permissibility theory, generates many permissibility results that we cannot accept – results that are, in some cases, highly counterintuitive and, in still others, conceptually at odds with one another. And that is so for both Exclusion Alpha and Exclusion Beta.

2.1.5 My guess is that the very real issues that Exclusion cannot seem to avoid, in either its Alpha or its Beta form, have led many theorists to think that Inclusion, however fantastic, *must be correct*.<sup>36</sup> The very aspect of Inclusion that makes it seem

---

<sup>35</sup> See Hare (2007), p. 503.

<sup>36</sup> Inclusion, in some form or another, is widely held. It is reflected in the total and average forms of consequentialism – Totalism and Averagism – as well as any form of Pluralism that considers the maximization of aggregate wellbeing among the goods that are to be taken into account in determining whether one outcome is better than another – and ultimately, perhaps, in determining the permissibility of the act that brings about the outcome rather than the other. (If betterness is not pertinent to permissibility, then I am unclear why we are investigating the former and – accordingly – unclear when a theory of betterness fails.)

One reason that Inclusion is so entrenched at present – though it has long commanded a good deal of interest and respect – perhaps has something to do with the *nonidentity problem*. An otherwise plausible permissibility theory that comports with Inclusion will consider the fact that a person *p* would be better off than a *nonidentical* person *q*, in a case where we cannot bring both *p* and *q* into existence and where, while *q*’s existence would be worth having, we cannot provide *q* with any additional wellbeing, to mean that bringing *q* into existence in place of *p* would be wrong. Parfit, most prominently, explores whether the nonidentity problem forces us to adopt an impersonal approach in *Reasons and Persons*. Parfit (1987), pp. 351–379. More recently, Caspar Hare explicitly rejects Exclusion, which he calls *Moral Actualism*, in both its forms. In turning to a “de dicto” solution for purposes of addressing the nonidentity problem, he can be viewed as accepting Inclusion. According to Hare, if conception could have been delayed so that “my third child” – *p* – would have been better off than “my third child” – *q* nonidentical to *p* – in fact is, then the failure to delay conception is wrong. How do we explain why that choice is wrong? It can’t be that the failure to delay is *better for q* – that *q* incurs a morally significant loss at, say, *w*<sub>2</sub>, where conception is not delayed – since *q* incurs no loss at all at *w*<sub>2</sub>; *q*’s wellbeing is there maximized. The wrongness Hare’s moral radar is picking up at this juncture would seem to have something to do with the stronger “moral pull” (my term, not his) the merely possible but better off *p* exerts on the agent than does the existing but less well off *q*. See Hare (2007), pp. 511–523. For critical discussion, see Wasserman (2008), pp. 529–535, and Parfit (1987), pp. 359–360.

It seems to me, however, that we should not be too concerned with the nonidentity problem in constructing our moral theory. It is not that I think that the choice of depletion or the risky policy, or the choice to enter into the slave child contract or to take the pleasure pill, is permissible, but rather that it is a fallacy to think that we cannot easily discern harm (in a comparative, intuitive sense of that term) in each of those cases. I thus argue that we can resolve the nonidentity problem without resorting to Inclusion. See part 2.2.5 below and Appendix B.

so counterintuitive on its face – the fact, that is, that it deems all losses of all persons at all worlds to matter morally, independent of who actually exists and independent of the act under scrutiny – is also what enables Inclusion to avoid the serious conceptual difficulties that plague Exclusion.

Now, there's no way around the fact that Inclusion generates highly counterintuitive permissibility results in some cases. But we've just noted that Exclusion, despite its air of commonsense, also generates highly counterintuitive permissibility results in some cases. As we shall see, the fact is that both Inclusion and Exclusion generate results that *only about half-way* conform to our intuitions.

So: neither Inclusion nor Exclusion perfectly conforms to our intuitions. Moreover, Exclusion faces deep conceptual difficulties that Inclusion manages to avoid. Are we, accordingly, compelled to make the leap from Exclusion to Inclusion?

For two reasons, we should first look. (A) There is too much at stake. Before we accept Inclusion, we need first to understand just how much we would be giving up. And (B) even if we agree that Exclusion is false, it is a mistake to think we are *compelled* to accept Inclusion. There are options beyond Exclusion and Inclusion, and we should not let a false dilemma force us to adopt Inclusion.

2.1.6 Let's take these two points in order. (A) Before we adopt Inclusion, we should understand fully what is at stake. Inclusion puts the *merely possible into a wildly implausible competition with those who do or will exist*. If we think the merely possible and *all their losses* matter morally in the way Inclusion suggests, we set up "as sure as fate" a "terrible spat" between merely possible persons on the one hand and existing and future persons on the other.

The air was littered, an hour or so,  
With bits of gingham and calico ...  
... the gingham dog and the calico cat  
Wallowed this way and tumbled that,  
Employing every tooth and claw  
In the awfulest way you ever saw ....<sup>37</sup>

Is this a "fate" that moral law can plausibly be understood to *require*? Can moral law *really not* make a distinction between your diminishing the wellbeing level of your own *existing* dog – tearing it to pieces, bit by bit – and your diminishing the wellbeing level of the *Calico Cat* by way of never bringing that cat into existence to begin with? It is all well and good to leave both parties in the "awfulest way" when they both do, or will, exist, we cannot avoid losses on behalf of either and we have distributed those losses between the two in whatever way it is that moral law requires. But the view that it is perfectly permissible for you to allow your own *flesh and blood, existing dog* to be left in the "awfulest way" so that the *Calico Cat* – or

---

Inclusion may also be at work in McMahan's own analysis of the Asymmetry and his treatment of abortion. See McMahan (2009), and (2006), pp. 625–655. I describe the Variabilist account of the Asymmetry in part 2.11 below and McMahan's treatment of abortion in Chapter 3 below. See also McMahan (2002), pp. 627 and 639–642.

<sup>37</sup> Field (1912).

the Cat Who Went to Heaven or Thomasina or any other *merely possible* cat – can be rescued from the *vast repository of individuals who will never exist at all* seems badly mistaken.

I believe that we in the end will be convinced that Exclusion in both its forms is fatally defective. But that fact is not going to make Inclusion seem *any less fantastic* to us than it does starting out. We are never going to say: Oh, *now* I see why Inclusion must be right. The one distinction that Inclusion insists we set aside is always going to seem to us one that no sound moral analysis can conceivably set aside: that one act imposes a loss on a *real, live, flesh and blood, sentient being* and the other a loss on, well, *nothing* that does or will *ever exist at all*. There *just is* an important moral distinction to be made between “making people happy” and “making happy people.”<sup>38</sup> In a way that can only be described as axiomatic, your actual dog *must* come before your merely possible cat.

The deficiencies in Inclusion are thus deep enough that rejecting Exclusion in favor of Inclusion does *not get us out of a bind*. It just *tightens the bind we are already in*. The purpose of this Chapter 2 is to *get us out*.

2.1.7 The second reason that any leap from Exclusion to Inclusion would be premature is that (B) there is a plausible middle ground between Exclusion and Inclusion. According to that middle ground, the merely possible do matter morally, just as you and I do, but they matter *variably* – and so do we. This middle ground could be called *Variable Exclusion* or – just as well – *Variable Inclusion*. I will settle for *Variabilism* here.

The strategy behind Variabilism is to identify, not which *people* matter morally and which do not, but rather which *losses* matter morally and which do not. We can then say that, for any person, whether existing, future or merely possible, some of the losses that person incurs matter morally and others do not.

As the cases will show, there is a pattern in when losses have moral significance and when they don't. But that pattern has nothing to do with *who* happens to incur the loss (whether, that is, the person incurring the loss *actually* exists, or *would* exist if the act under scrutiny were performed, or etc.). Instead, what imbues some losses but not others with moral significance – what makes some losses such that, in calculating what we ought to do, we *must* take them into account, and makes other losses such that we *must not* take them into account – is just *where* those losses are incurred in relation to the person who incurs them. Incurred in a circumstance, or possible future or *world*, where a person does or will exist, a loss will have *full moral significance*. Incurred anywhere else, a loss will have *no moral significance whatsoever*, not even the slightest bit.

The availability of Variabilism thus shows that, if, having established that Exclusion is false, we *do* infer that Inclusion is true – if we do make the leap from Exclusion to Inclusion – it isn't because we *must* but instead only because we mistakenly *think* that we must.

---

<sup>38</sup> Narveson (1976), p. 73.

## 2.2 Preliminaries – A Maximizing Account of Loss; the Loss of Never Existing; the Loss of Death; the Otherwise Plausible Permissibility Theory

2.2.1 *Loss*. Some preliminaries are in order. For purposes here, I adopt a *maximizing* account of harm, or *loss*. I believe that my account of *loss* captures one ordinary concept of *loss* (or *harm*) that we have. For purposes here, however, the important thing is that the term *loss* is shorthand for a certain *betterness* relation. Thus: to say a person incurs a *loss* at one world under one act is just to say that an alternative act performed at an alternative world is *better for* that person than is the one act performed at the one world.

In other words: a person incurs a *loss* whenever agents (by act or omission) create less wellbeing for that person when agents *could have created more wellbeing* for that very *same person*. That is so, even if it is also true that those same agents *would* never in a million years have been willing to create that additional wellbeing for that person.

One feature of the maximizing account of loss will be especially important going forward – in particular, when we begin to compare an act *a* performed at a world *w* against an act *a'* performed at a world *w'* for purposes of determining the permissibility of act *a*. On my account, a person can incur a loss at *w* *not just* in virtue of how that person fares at *w* relative to how that same person fares at *w'*, *but also* in virtue of how that person fares at *some third world w''* under some alternative act *a''*. This feature will mean that the loss incurred under *a'* at *w'* *relative to w''* – if that loss is deemed morally significant – can easily count against *a'*, notwithstanding the fact that the person incurs no loss at all at *w'* *relative to w* (i.e. is no worse off at *w'* than that same person is at *w*).

The maximizing account of loss is inherently *comparative*. Incurring a loss is a matter of being made worse off. And we determine whether a person is made worse off by comparing how well off that person is in one circumstance, or possible future or world, with how well off that same person is in another.

*Temporal* and *counterfactual* (or *but for*) accounts of harm, or loss, also operate by comparing wellbeing levels for the same person from one circumstance to another. But such accounts notoriously undercount loss. The temporal account considers worsening over time to be a necessary condition for loss. But clearly a person can be made better off, or at least not worse off, from one moment to the next and *still incur a loss*.<sup>39</sup> The reason we think that the surgical team who has done a sloppy job restoring your vision – you now can see, but not well – has caused you to incur a *loss* is that we understand that the team *could have done more for you* than it has.<sup>40</sup> Whether the team was *obligated*, morally or legally, to have acted with greater care on your behalf – whether, that is, you have been *wronged* – is a different question. That you are better off does not mean you have not been harmed or that you have not incurred a loss.

According to the counterfactual account, an act imposes a loss on a person *only if*, had agents not performed that act (*but for* that act), that person *would have been*

<sup>39</sup> Feinberg (1988).

<sup>40</sup> This example is based on Hanser (2009).

better off. But, clearly, I have caused you to incur a loss when I shoot you in the arm even if I was so angry that, had I not shot you in the arm, I *would have* shot you in the heart. In a way, I have *saved your life* by shooting you in the arm. But still you incur a *loss*. According to the maximizing account of loss I propose here, it is a simple thing to explain why that is so: I *could have* refrained from shooting you altogether; I *could have* done more for you than shoot you in the arm. The difference between what I did and what I could have done counts as a *loss*.

The maximizing account recognizes, in other words, that a person can (A) incur no loss at all between one world and a second world – for example, the world we count, for purposes of evaluating the counterfactual, as the “closest” world at which the act under scrutiny is not performed – but (B) still incur a loss between the one world and some third – a world agents *could* have brought about, independent of what they *would* have done; a world that was *accessible* to agents at the time just prior to performance.<sup>41</sup>

That the maximizing account of loss is broader than some of its competitors does not mean that we necessarily end up with an unduly broad concept of *wrongdoing*. After all, once we identify loss, we must then determine whether that particular loss is *morally significant*. If we determine that it is, we must then appeal to the otherwise plausible permissibility theory to determine whether a *wrong* has been done.

At the same time, the account of loss that I propose is narrower in certain respects than some other accounts. For example, the fact that it is inherently *comparative* narrows its scope. On my account, a person incurs a loss only if that person is worse off than he or she is relative to some baseline.<sup>42</sup> In contrast, a *non-comparative* approach defines loss by reference to a list of specific bad effects (e.g. substantial pain or impairment) or a threshold (e.g. a minimally decent wellbeing level).<sup>43</sup> But it is unclear to me that a person whose wellbeing has been maximized may also have incurred a *loss*. It is unclear to me that the vaccination – or the kidney transplant – that imposes pain has also imposed a *loss* or a *harm* in any sense that is relevant to the issue of wrongdoing. A maximizing account of loss easily explains why the patient incurs no loss in such cases: when wellbeing has been maximized, there is no *loss*, even if there is pain.

On my account of loss, determining whether an act imposes a loss when performed at a given world obviously supposes that any burdens imposed by that act are to be netted against the wellbeing that is created for that person by that act. This point is somewhat controversial. Where, however, we limit our focus to what a given act does to a given person at a given world, netting seems in order. It is netting

---

<sup>41</sup> I take it for granted that a person’s incurring a *loss* at a world requires a cross-world *identity* but does *not* require that person’s having an alternate better off “self,” in Velleman’s sense, at an alternate world (other than to the extent that there being a “self” there coincides with the person who incurs the loss here being *identical* to a better off person there). I thus set Velleman’s skeptical arguments aside for purposes here. But see Velleman (2008).

<sup>42</sup> Feinberg (1988).

<sup>43</sup> See Harman (2009), p. 139, and Harman (2000), p. 93, for discussion of a list approach to harm. See also Shiffrin (1999), p. 123. For discussion of the relevance of the minimally decent life, see Steinbock (2009), p. 164.



burdens against benefits that tells us, for example, that the vaccination, despite the pain, does not impose a loss.<sup>44</sup>

In still another way, the account of loss that I adopt here is narrow. On my account, *loss* does not include damage inflicted by nature unless agents could have interceded to avoid that damage. I think there is a familiar sense of the term *loss* according to which some of the deepest losses we will ever incur are thanks to nature. So connecting loss with whether agents could have done more for a person than they have done may seem artificial. On the other hand, there is no reason to think that my account of *man-made* loss cannot coexist with a separate account of *natural* loss. But we have no reason for undertaking that latter account here.

**2.2.2 The Loss of Never Existing.** I accept for purposes here that we can cogently compare a person's wellbeing level at a world where that person does or will exist against a person's wellbeing level at a world where that same person never exists.<sup>45</sup> Thus, as noted earlier, it is at least plausible that a person accrues *no wellbeing at all* – no benefits, no burdens, no goods, no “bads” – at any world at which that person never exists. We can say that persons who never exist at a world have a zero wellbeing level at that world.

We can then say that an agent's leaving a person out of existence altogether at a world often creates less wellbeing for that person when the agent could have created more wellbeing for that same person by bringing that person into an existence worth having. It often makes things worse for that person when things might have been made better for that same person. Accordingly, leaving a person out of existence altogether at a world often causes that person to incur a *loss* at that world.

Similarly, bringing a person into existence often creates more wellbeing for that person when agents could have created less wellbeing for that person by leaving

---

<sup>44</sup> See Benatar (2006), pp. 18–22, for the argument that for purposes of determining harm burdens are not to be netted against benefits.

<sup>45</sup> I have elsewhere argued that comparisons between a person's wellbeing level at a world where that person exists and a person's wellbeing level at a world where that person never exists are cogent. See Roberts (2003a). Since we are here investigating (among others) Inclusion, according to which the merely possible *do* matter morally and the losses they incur when we leave them out of existence altogether (or bring them into an existence that is *less* than worth having) *do* bear on the permissibility of the acts we perform, it can only add clarity to talk openly about such losses. Assigning a positive number to the existence worth having and a negative number to the existence less than worth having (even if those numbers have their ordinal value only) and zero to the state of never having existed at all is also useful. See too Nils Holtug (2001), pp. 361–384, and Fred Feldman (1991), p. 210 (“I stipulate that if *s* fails to exist at *w*, then  $V(s, w) = 0$ .”). But see also Krister Bykvist (2007), pp. 339–345; John Broome (1999), p. 168; McMahan (1999), p. 168; Clark Wolf (1997), pp. 108–109; David Heyd (1992), pp. 30–33; Cynthia Cohen (1996), p. 22; and Partha Dasgupta (1995), p. 383 (“It makes no sense to attribute a degree of well-being, low or high or nil, to the ‘state of not being born.’ Non-existence is like nothing for us, not even a very long night, because there is no *us* to imagine upon. One can't be asked what it would be like to experience one's own nonexistence, for there is no *subject* of experience in non-existence.”). More recently, Broome, citing Wlodek Rabinowicz, has suggested that, even if it is not cogent to say that a person is better (or worse) off “in” a world in which that person never exists, it may still be “analytically possible” to say that a world in which a person never exists is better (or worse) “for” that person. Broome (2004), p. 63. See also note 32 above.



that person out of existence altogether. Bringing a person into an existence worth having thus *avoids a loss* on behalf of that person.

But we must consider the unusual cases as well. We must recognize, that is, that leaving a person out of existence altogether sometimes creates more wellbeing for that person than bringing that person into a genuinely *wrongful life* – an existence that is *less* than worth having – does. In such cases, leaving the person out of existence altogether is what *avoids a loss* on behalf of that person. And bringing the person into existence is what *imposes a loss* on that person.

An implication of this way of valuing the plight of never having existed at all is that a person may incur distinct losses at a given world at which that person never exists relative to the distinct alternative existences he or she might have been brought into. *Anyone's* life can be made less than worth living; *wrongful life* is the case where that outcome cannot be avoided if one is to exist at all.

**2.2.3 The Loss of Death.** It is plausible that, once someone dies, we can no longer act in a way that makes things better or worse for that person. It is plausible that additional wellbeing or illbeing cannot continue to accrue to a person *after* that person ceases to exist. If that is correct, then it is misleading to talk as though agents can perform acts that cause persons who once existed but no longer exist to incur losses. In particular, it is misleading to suggest that the losses they incur when still alive can be made up for by post-existence gains. Accordingly, for purposes here, I avoid any reference to acts that impose losses on persons who *have existed*. The standard reference, instead, is to losses that are incurred by a person who *does or will exist* at the world at which the loss is incurred.

Of course, it would be a problem if an account formulated in this way meant that death was not itself a loss. But it doesn't. Even given that the incurrence of a post-existence loss is not possible, that *last moment of existence* – like any other – is one during which losses *can* be incurred. That the *effect* of a particular act – the effect being that the person *no longer exists* – goes *beyond* that last moment of existence *does not mean* that that person does not *incur* the loss at that last moment. That the *effect* of death materializes only *after* the person goes out of existence does not mean that the *loss* is not incurred *before* that time.<sup>46</sup>

---

<sup>46</sup> In other words, we thus can discern a loss, when (i) agents act at t1 in a way that makes t2 the last moment of a person's existence and that accordingly generates no future at all for that person at t3 and (ii) those same agents could have instead acted at t1 in a way that would have meant that t2 is *not* the last moment of a person's existence and that would accordingly have generated a future worth having for that person at t3. They have acted in a way, at t1, that imposes on that person – causes, that is, that person to incur – a loss at t2 – a loss that itself is incurred at t2 but doesn't materialize until t3. It isn't, in other words, the *death* that imposes the loss. It's the *act* that imposes the loss – a loss that may be incurred, for example, during that last moment of existence and just prior to death.

By implication, I am here accepting a comparative account of the loss of death. Feldman also describes a comparative account of the loss of death, but the comparison he would make is between the future as it in fact unfolds and the future that unfolds at the nearest world where the person does not die (the future that would have unfolded had the person not died). Feldman (1992), p. 216–217. See also Feldman (1994). But see part 2.2.1 above (discussion of difficulties inherent in counterfactual account of loss).

This means that the death of someone else is not just a loss *we* incur. It's a loss *they* incur as well. They just don't *experience* it, and we do.

But the loss itself is important, not just the experience of the loss. The experience of the loss should be understood as a separate loss. We aren't just grieving for ourselves when someone dies. We aren't just grieving for ourselves and a loss we have incurred. We are also grieving for the person who has died. We are grieving, that is, for the fact that that person has incurred a loss.

*2.2.4 The Otherwise Plausible Permissibility Theory.* The competing rules I examine here are not *moral theories* in a traditional sense. They don't, on their own, decide just which acts are permissible and which are not. Instead, they decide just which *losses* have moral significance and which do not. Such rules – *loss rules* – generate different, and competing, collections of *moral data* on the basis of which an independent and otherwise plausible permissibility theory can then generate different, and competing, permissibility results.

Clearly, if the testing that we do of Exclusion Alpha against Exclusion Beta, Exclusion Beta against Inclusion, Inclusion against Variabilism, and so on, is to be meaningful, the theory that we rely on to generate the permissibility results we use for that purpose must be *otherwise plausible*. The theory must be plausible other than in respect of the losses deemed morally significant by whatever loss rule it is we happen to be testing. Without that assurance, we won't have a clue whether any of the many transparently false results we obtain in what follows are to be ascribed to the loss rule we are then testing or to the permissibility theory.

We thus will need to say something about the content of any such *otherwise plausible permissibility theory*. We won't agree, starting out, just which losses are morally significant and which are not. But setting that issue aside, we will find it easy to agree, I think, on a handful of principles that simply *connect* (A) the moral data they are given by whatever loss rule we happen to be testing, with (B) specific permissibility results, for some range of cases that is of interest to us.

And a handful of principles is all we need. For purposes here, we don't need a *complete* permissibility theory. A rudimentary theory will do. Our needs are modest since the test cases will be simple.

We can then begin with a Pareto principle of sorts.

*Otherwise Plausible Permissibility Principle 1 (OPPP1):*

An act *a* performed at a world *w* is wrong *if*

- (i) Act *a* imposes a morally significant loss on a person *p* at *w*;
- (ii) An act *a'* at an accessible world *w'* performed in lieu of *a* at *w* is better for *p* at *w'* than *a* is for *p* at *w*; and
- (iii) Either (a) *a'* imposes no morally significant loss on any person *q* at *w'* other than perhaps *p*, or (b) if *a'* imposes a morally significant loss on a person *q* at *w'*, then *a* imposes a morally significant loss on *q* at *w*, and *a'* is at least as good for *q* at *w'* as *a* is for *q* at *w*.

The Pareto idea – roughly put – is that if agents can make things better for at least one person and worse for none and fail to do so, then what they have done is wrong. OPPP1 restates that idea so that certain critical questions are explicitly left open.

OPPP1 doesn't leave *every* question open. The maximizing account of loss I have adopted here is meant to be incorporated into OPPP1. Thus, for purposes here, it is settled that a loss is incurred by a person when that person is left out of an existence worth having. What clause (i) of OPPP1 leaves open is the moral significance of that loss.

That OPPP1 incorporates the maximizing account of loss settles still another important point. Clause (iii) is potentially failed when  $q$  incurs a loss at  $w'$ . What is important for purposes here is that loss itself is equally well established by comparing how well off  $q$  is in  $w'$  against how well off  $q$  is in  $w$  and by comparing how well off  $q$  is in  $w'$  against how well off  $q$  is in some third world  $w''$ .

Finally, just as a loss is imposed when agents leave a person out of an existence worth having, bringing a person into an existence worth having can be a way of making things better for that person for purposes of clause (ii).<sup>47</sup>

OPPP1 thus takes as settled certain issues about loss and what counts as making things better for a person. But OPPP1 leaves open whether those losses count as *morally significant losses*. It leaves open whether (A) the loss  $p$  incurs at  $w$  in the case where  $p$  never exists at  $w$  and (B) the loss  $q$  incurs at  $w'$  in the case where  $q$  never exists at  $w'$  count as *morally significant losses*. Thus, if leaving  $p$  out of an existence worth having at  $w$  is a way of imposing a morally significant loss on  $p$  at  $w$ , then OPPP1 may on that basis (and depending on other facts) imply that  $a$  is wrong – that it is wrong *not* to bring  $p$  into existence. But if it doesn't, OPPP1 will have no application at all. Some of the loss rules we examine in what follows take one position on that issue; some the other. But that issue is part of what is at stake here. So the otherwise plausible permissibility theory – the theory that determines permissibility – must, if it is to be the sort of widely accepted and uncontroversial theory that we can use for purposes of testing, avoid resolving that particular issue in advance.

What distinguishes OPPP1 from some rough formulations of the Pareto idea is this: when we compare an act performed at one world against an act performed at another world for purposes of evaluating the one act, we do not just ask whether things have been made worse for any person at that other world than they are at the one. We ask a broader question: whether a person has been made to incur a morally significant loss at that other world. And that is a matter *not just* of how that person fares at the other world relative to the one, *but also* of how that person fares at any *third* world relative to the one.

Whether any such loss itself has moral significance is something OPPP1 leaves open. Specifically, clause (iii) of OPPP1 leaves open the issue of whether the loss  $q$  incurs at  $w'$  *compared to how well  $q$  fares at some third world  $w''$*  counts as a morally significant loss. Or do we say instead that  $q$ 's loss at  $w'$  counts as morally significant, for purposes of evaluating  $a$ , *only if  $q$  incurs a loss at  $w'$  compared to how  $q$  fares at  $w$* ?

---

<sup>47</sup> See parts 2.2.1 and 2.2.2 above.

This issue is potentially important in any additional person context. For, if  $q$  doesn't exist at all at  $w$ , then  $q$  will not incur a loss at  $w'$  *compared to how well  $q$  fares at  $w$*  so long as  $q$  has an existence worth having at  $w'$ . If *that* is the clause (iii) test, then it is easily satisfied. But there is nothing in OPPP1 that indicates that that narrow reading is in order. Consistent with OPPP1, clause (iii) may be far more stringent. That  $q$  incurs no loss at  $w'$  compared to how  $q$  fares at  $w$  doesn't at all imply that  $q$  incurs *no morally significant loss at  $w'$  at all*.

And it is critical that OPPP1 avoid taking any position on that issue. Whether the loss  $q$  incurs at  $w'$  relative to some third world  $w''$  counts as morally significant is part of what is at stake here. Some of the loss rules that we will examine will say that it is, and some that it isn't. It is, accordingly, critical that that issue remain open in the context of the otherwise plausible permissibility theory.

Still another question OPPP1 leaves open is its relation to the person-affecting, or person-based, intuition. OPPP1 provides only a sufficient, and not a necessary, condition for wrongdoing. That means that OPPP1 leaves open the question of whether  $a$  is wrong in a case where agents *do not have the alternative of making things better for the identical person  $p$*  who incurs a morally significant loss at  $w$ . Consistent with OPPP1, agents have done something wrong – not by implication from OPPP1, but rather by implication from some other sufficient condition on wrongdoing – when they (1) bring  $p$  into an existence that is unavoidably flawed but worth having and (2) had the alternative of bringing a better off but *nonidentical* person  $q$  into existence in place of  $p$ .

In other words: OPPP1 does not imply the person-based intuition, *narrowly construed*.<sup>48</sup> And that is a good thing, since the truth of the person-based intuition, narrowly construed, is also part of what is at stake here. The loss rules we examine in follows, in other words, take very different positions on just that issue. So it is an issue our otherwise plausible permissibility theory must itself leave open.

In a moment, we will turn to a quite different necessary condition on wrongdoing, a form of the person-based intuition that we *can* include in the otherwise plausible permissibility theory by the virtue of the fact that it leaves open whether it is to be narrowly or more broadly construed. But first there is more to be said about OPPP1.

The skeletal OPPP1 is just one Pareto principle. We examine two others in Chapter 3. The first is the nonstandard Pareto principle sometimes called *Pareto Plus*.<sup>49</sup> The rough idea there is that it is wrong not to bring a person into an existence worth having when agents have the alternative of doing that without making things worse for any person. Pareto Plus is, however, obviously too controversial to be included here as part of an otherwise plausible permissibility theory. Its truth, too, is at stake; and we accordingly are not going to assume its truth starting out. To put the point another way: Pareto Plus is *what we get* when we combine Inclusion with the otherwise plausible permissibility theory (which itself includes OPPP1).

<sup>48</sup> See note 17 above (relation between person-based approach and Moral Actualism).

<sup>49</sup> Dasgupta (1995), pp. 382–83. The principle Dasgupta calls *Pareto Plus* is stronger than what I call *Pareto Plus* here. See note 142 below.

The second Pareto principle that we will examine in Chapter 3 is more standard. It is limited to the case where the choices to be compared involve identical populations – when they are, that is, “same-people” choices. OPPP1, in contrast, is not limited in that way. Accordingly, in cases in which leaving one person out of existence makes things better for still others, OPPP1 may *require* agents *not* to bring that person into existence to begin with, depending on the loss rule we in the end adopt. But that result is nothing we are going to build into the otherwise plausible permissibility theory itself. Whether or not that result is correct is part of what is at stake here. It is what we get – and *Pareto Minus* would be good name for it – when we combine Exclusion, or Variabilism, with OPPP1.

A result we will be able to obtain from OPPP1 under some, but not all, of the loss rules we examine in what follows involves the case where the life is both unavoidably less than worth living and actual. The life is, that is, *genuinely wrongful* and the person who incurs it is someone who exists at what is uniquely the *actual* world. The maximizing account of loss immediately implies that that person incurs a loss. Inclusion, Exclusion and Variabilism all agree that that loss is morally significant. Where leaving that person out of existence altogether would not have created a morally significant loss for anyone at all, OPPP1 easily implies that bringing that person into existence is wrong.

A second component of the otherwise plausible permissibility theory is a version of the person-based intuition. It provides not a sufficient but rather a *necessary* condition on wrongdoing – by implication, a *sufficient* condition for permissibility.

*Otherwise Plausible Permissibility Principle 2 (OPPP2):*

An act *a* performed at a world *w* is wrong *only if* act *a* imposes a morally significant loss on a person *p* at *w*.

Wrongdoing requires, in other words, that someone or another – some existing person or some future person or, consistent with OPPP2, some *merely possible* person – incur a morally significant loss.

Narrowly construed, OPPP2 will identify quite a range of acts as permissible. But it's nonetheless suitable for inclusion in the otherwise plausible permissibility theory because it is consistent with both a *narrow* form and an alternate *very wide* form of the person-based intuition.<sup>50</sup> OPPP2 thus should not be considered *at all* controversial for purposes here. After all, the class of persons we quantify over in OPPP2 *includes* the merely possible. OPPP2 thus leaves open the question whether its necessary condition is satisfied when the loss that is incurred is incurred by a person who never exists at all. As long as the loss is morally significant, the necessary condition will be satisfied. Some of the loss rules we will be examining say that it is, and others that it isn't.

Neither of the two principles we have included in the otherwise plausible permissibility theory so far addresses overt tradeoff scenarios. In part, that is so because the

---

<sup>50</sup>This latter way of construing OPPP2 may be equivalent to what Parfit calls the “wide person-affecting approach.” Parfit (1987), pp. 396–401.

whole topic of how tradeoffs are to be resolved is inherently controversial. Nonetheless, the following principle surely would command wide agreement. The idea behind this principle is that an act is permissible if each alternative act involves a *merely reversing change*.<sup>51</sup> Thus:

*Otherwise Plausible Permissibility Principle 3 (OPPP3):*

An act *a* performed at a world *w* is permissible *if*

- (i) Act *a* imposes a morally significant loss on a person *p* at *w*;
- (ii) Each alternative to *a* – that is, each *a'* performed at an accessible world *w'* in lieu of *a* performed at *w* – imposes a morally significant loss on a person *q* (who may but need not be identical to *p*);
- (iii) *a* performed at *w* is at least as good for *p* as *a'* is for *q* at *w'*; and
- (iv) Either (a) *a* imposes no morally significant loss on any person *r* at *w* other than *p*, or (b) if *a* imposes a morally significant loss on any person *r* at *w* other than *p*, then *a'* imposes a morally significant loss on *r* at *w'* and *a* performed at *w* is at least as good for *r* as *a'* is for *r* at *w'*.

OPPP3 proposes a sufficient condition on permissibility. Suppose that agents have *just two* alternatives *a* and *a'*, and the loss *q* incurs under *a'* is at least as great and deep as the loss *p* incurs under *a*, and the acts otherwise impose no morally significant losses at all. OPPP3 then implies that *a* is permissible. Moreover, when the two acts both impose morally significant losses, and those losses are themselves equally great and equally deep, then *it doesn't matter* which act agents perform. *Both* acts are permissible; the distinctions between them are, in effect, deemed immaterial.

Clause (i) limits the principle to the case where the act imposes a morally significant loss. If the act under scrutiny imposes no morally significant loss on anyone at all, then we need not bother with OPPP3. We can just appeal to OPPP2 directly to determine that that act is permissible.

OPPP3, like the other principles that qualify for the otherwise plausible permissibility theory, leaves some important issues open. Thus, clause (i) leaves open whether *p* must exist at *w* to incur a morally significant loss at *w*. Similarly, clause (ii) leaves open whether *q* must exist at *w'* to incur a morally significant loss at *w'*. Some of the loss rules we examine here say that that loss is morally significant even if *q* does not exist at *w'* and some say that it's not. If the loss of having been left out of existence altogether *does* count as morally significant, then two acts in some cases will be considered a *wash* even though one imposes a loss on an existing person and the other on a person who never exists at all. If that loss does not count as morally significant, OPPP3 will have no application at all.

One last point. The rationale for clause (iv) is just to insure that act *a* is not flowed in some way that act *a'* is not and that has nothing to do with *p* or *q*. Clause (iv) thus insures that we will not conclude, on the basis of what *a* does to still

---

<sup>51</sup> See generally Peter Vallentyne (2000), pp. 1–19.

others, that  $a$  is wrong. The two acts are, in other words, otherwise morally indistinguishable – there is otherwise *nothing about the alternative act* that might make us think that the one act is itself wrong.

A final principle declares an act wrong if it imposes a morally significant loss on one person and agents could have avoided that loss on behalf of that person by imposing a less severe morally significant loss on at most one other person. This principle, like OPPP3, addresses a handful of very simple tradeoff scenarios.

*Otherwise Plausible Permissibility Principle 4 (OPPP4):*

An act  $a$  performed at a world  $w$  is wrong *if*

- (i) Act  $a$  imposes a morally significant loss on a person  $p$  at  $w$ ;
- (ii) An act  $a'$  performed at an accessible world  $w'$  in lieu of  $a$  performed at  $w$  is better for  $p$  at  $w'$  than  $a$  is for  $p$  at  $w$ ;
- (iii) If  $a'$  performed at  $w'$  imposes a morally significant loss on any person  $q$  at  $w'$ , then that morally significant loss is not as great and not as deep as the morally significant loss that  $a$  performed at  $w$  imposes on  $p$ ; and
- (iv) Either (a) at most one person –  $q$  – incurs a morally significant loss under  $a'$  at  $w'$  or (b) if  $a'$  imposes a morally significant loss on any other person  $r$  at  $w'$ , then  $a'$  is at least as good for  $r$  at  $w'$  as  $a$  is for  $r$  at  $w$ .

Thus, as between two otherwise morally indistinguishable acts, OPPP4 requires agents to impose the lesser and shallower morally significant loss on one person rather than the greater and deeper morally significant loss on someone else. It is left open, as before, whether failing to bring someone into an existence worth having imposes a morally significant loss.

OPPP4 can be viewed as an extension of OPPP1. Whereas the usual Pareto idea applies when agents have the alternative of making things better for a *particular person* without making things worse for anyone else, OPPP4 applies when agents have the option of making things better for  $p$  than they are for a *distinct person*  $q$ .

OPPP4 governs *very* simple tradeoff scenarios. We actually will not need OPPP4 until we turn to the problem of late abortion in Chapter 3. But we do face a handful of somewhat more difficult tradeoff scenarios in this Chapter 2. And in that connection we will consider what sorts of tradeoff principles seem plausible. But those principles, unlike OPPP4, are just a bit too controversial to be included in the otherwise plausible permissibility theory.

More generally, other principles will be considered as needed, with any controversies noted as we go. The general assumption is that principles that become part of the otherwise plausible permissibility theory will take a *comparative* approach to determining permissibility. Just like OPPP1–OPPP4, they will evaluate the permissibility of a given act by reference to *just how dreadful the alternatives to that act are*. If all the alternatives to a given act are *just as dreadful* as the one act, then the one act is permissible. If the one act is *more dreadful* than some alternative act, then the one act is wrong. But by *more dreadful* here



I do not mean that one alternative act generates less wellbeing *in the aggregate* than the other.<sup>52</sup> Rather, I am just recognizing in an informal way that alternative acts often have certain features that tell us something about the permissibility of any one act.

In addition, we can freely appeal to various principles that we can describe as *standard deontic axioms*. Such axioms plausibly include: if an act is obligatory, it is permissible; if an act is obligatory, none of its alternatives are permissible; and if each alternative act is wrong, then any one act must be permissible.<sup>53</sup> Moreover, if the otherwise plausible permissibility theory happens to be relevantly *complete*, and if no alternative exists for agents that, according to that theory, shows that the one act is *wrong* – if, that is, *each such alternative is at least as dreadful as the one* – then we will have no basis for thinking that the one act *is* wrong. And we can then say that that act is permissible.

We can think of such deontic axioms as part of any otherwise plausible permissibility theory. But we need not do so. It doesn't matter. If they are axioms, we can count on them being there for us in any event. And – of course – we must make sure that our otherwise plausible permissibility theory complies with them.

This discussion shows that, while the maximizing account of loss I adopt for purposes here is broad, we will not necessarily end up with an unduly broad concept of *wrongdoing*. Agents often face tradeoffs. They often must choose which of two persons shall incur a loss, and the loss rules we examine in what follows often deem *both* such prospective losses morally significant. But it won't follow that the acts that impose those losses will *both* be wrong. Instead, as long as the tradeoff between the persons in conflict is made in an *appropriate* way – is made, that is, in the way that the *otherwise plausible permissibility theory* directs – what the agent has done will be deemed permissible.

**2.2.5 The Nonidentity Problem.** It seems indisputable that the choices we make today regarding the *wellbeing levels* to be accorded to future persons can have a very significant impact on the *identities* of those future persons. This fact raises the specter of the *nonidentity problem*. Some theorists might consider it a mistake to rely so heavily on an account of loss that requires for loss that agents create less wellbeing for a person when they could have created more wellbeing for that *same identical* person. They might think that the nonidentity problem means that many cases in which we *want* to say (and naively think we *can* say and *need* to be able to say, if we want to make loss so central to an account of wrongdoing) that a loss *bears on* the permissibility of a given act are cases that, on closer inspection, do not,

---

<sup>52</sup> Thus we do not assume that the dreadfulness of a given alternative is just a matter of the aggregate wellbeing, or amalgamated overall good, we see under that alternative. To assume that would be to assume that the loss of leaving a person out of an existence worth having is itself a *morally significant* loss, and that is not something that the otherwise plausible permissibility theory should make any assumptions about one way or the other. That issue, rather, is what is at stake in this book.

<sup>53</sup> For an account of the fundamentals of deontic logic, see Paul McNamara (2006).

according to the maximizing account I adopt here, involve loss at all. We can bring a *nonidentical* person into a better existence, but we cannot (or so the nonidentity problem asserts) bring that *very same person* into a better existence.

For two reasons I only briefly address the nonidentity problem here. First, I have discussed it in detail elsewhere.<sup>54</sup> And, second, I believe that the nonidentity problem in fact undermines *nothing* we will want or need to say here.

I outline my response to the nonidentity problem in Appendix B below. Very briefly here, I begin by sorting the problems that are collected together under the heading of “nonidentity” into their various problem *types*. Perhaps the most challenging type of nonidentity problem includes Parfit’s depletion and risky policy examples, Kavka’s slave child and pleasure pill cases and cases involving historical injustice. Even if a better-for-*p* alternative world exists and is accessible to agents, what I call the *can’t-expect-better* problem (correctly) notes that the *chances* that agents would have been able to bring about that particular world (or one relevantly like it; one that includes a better off *p*) in place of a world in which *p* never exists at all are miniscule. But isn’t it *better for p* to have the flawed existence than to have the miniscule chance of an unflawed existence? And doesn’t that mean, under a maximizing account of loss, that *p* has incurred no loss at all? But isn’t the choice under scrutiny nonetheless clearly wrong?

A second type of nonidentity problem – the *can’t-do-better* problem – applies in the case where there is no world that is both better-for-*p* and accessible to agents. A good example of this second type of problem is the case where a child is born with a serious genetic disorder we do not yet have the ability to correct or to completely palliate, and the child’s wellbeing level is unavoidably diminished compared to – say – the norm. Isn’t it nonetheless better from that child’s own perspective to have the flawed existence rather than none at all? (We assume that the child’s life remains worth living – that we are not dealing here with the *genuinely wrongful* life.) And doesn’t that mean that the child incurs no loss? But, when agents had the alternative of bringing a nonidentical but better off child into existence in place of the one, isn’t the choice to bring the impaired child into existence clearly wrong?

Having divided the cases, I then argue that we can resolve “the nonidentity problem” – that is, that we can resolve *each type* of nonidentity problem on its own terms – so long as we take them *one-by-one* and *not both at the same time*. And I argue that we can do so *without* embracing Inclusion.

Thus, the can’t-expect-better version of the nonidentity problem is, I argue, the product of what I elsewhere call the *nonidentity fallacy*.<sup>55</sup> I ought to have said *fallacies*. It is a fallacy to think we can infer that no loss is imposed on a particular person from the fact that the *actual* wellbeing level generated for a person by one act is greater than the *expected* wellbeing level generated for that same person by

<sup>54</sup> See Roberts (2009a, c, 2007).

<sup>55</sup> See Roberts (2009c, 2007).

the other act. When we rewrite the argument, so that we are just comparing one *expected* wellbeing level against another *expected* wellbeing level, we happen onto still another fallacy: the fallacy of thinking that just because a given wrong act has been performed, and the future has unfolded in a way that includes a particular child, the probability, at that critical time just before choice, is more than miniscule that that child will exist, given that choice. The fact is that what we have is a wash: the odds against any particular child coming into existence are *very great*, whether the wrong act is performed or a better act is performed in its place.

The can't-do-better problem involves the rare case where the choice to bring an unavoidably impaired child into existence maximizes wellbeing for each and every person who does or will exist. We may agree, depending on which loss rule we at the end of the day adopt, that such cases involve no morally significant loss at all – none incurred by those who do or will exist, and none incurred by any merely possible person. But it seems to me that we should also think, about those same cases, that it is not clear that a wrong has been done.

More typically, of course, the choice to bring an unavoidably impaired child into existence involves losses that will count as morally significant whatever loss rule we at the end of the day adopt. We can, that is, easily find what we can think of not as a *procreative* loss but rather as a *distributive* loss – though still a bona fide *loss*. That loss can just as easily be incurred by the genetically impaired child or by that child's already-existing or later siblings or by still others. Depending on the facts, that loss can then easily support the result that what the agents have done is wrong.

## 2.3 The Basic Case

2.3.1 The *Basic Case* underlines what we like about Exclusion (Case 2.1). It parallels the case of the “terrible spat” between the existing dog (gingham or not, as long as it qualifies as a *person*) and the merely possible cat.

Suppose we must choose between acts  $a_1$  and  $a_2$ , where  $a_1$  can be counted on<sup>56</sup> to bring about:

---

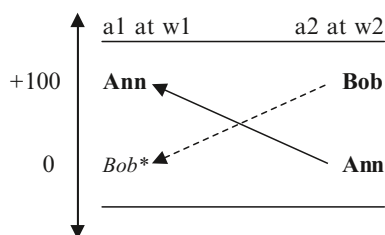
<sup>56</sup>In general, we have little control over what possible future, or world, in all its specificity, our choice will produce. For one thing, there may be no settled matter of fact that there is any one particular way the future will unfold if we happen to choose  $a_1$  in place of  $a_2$ . For another, there may be many different ways of realizing  $a_1$  itself, and those different ways might themselves trigger different futures. Still, in some cases, we can control whether one particular outcome or another is *part of* whatever world it is at which we perform one particular act rather than another. Let's suppose that we are in that kind of case here.

A person Ann's having an existence that is well worth having, and a possible person Bob's never coming into existence at all and for that reason accruing an overall wellbeing level of exactly zero.<sup>57</sup>

And a2 can be counted on to bring about:

Ann's coming into existence and having an overall wellbeing level of exactly zero, and Bob's coming into an existence that is just as worth having as Ann's is under a1.

We can diagram the case as follows. Where (i) natural numbers represent overall wellbeing levels<sup>58</sup>, (ii) a1 and a2, which exhaust our alternatives, are performed at (arbitrary) worlds w1 and w2 where they respectively bring about the outcomes just described, (iii) boldface means that a specified person *does or will exist* at the specified world and italics, followed by the asterisk, that the specified person *never exists at all* at the specified world; and (iv) no one other than Ann and Bob have anything at stake in whether a1 or a2 is performed:



Case 2.1 Basic Case

a2 makes things much worse for Ann than a1 does. She, accordingly, incurs a loss under a2 but not a1. There is, moreover, at least a case to be made – a case that, as we have seen, is taken for granted for purposes here – that a1 makes things much worse for Bob than a2 does, and that Bob, accordingly, incurs a loss under a1 but

<sup>57</sup>That Bob's overall wellbeing level is zero at any world where Bob never exists at all is an assumption that I make for purposes here. More generally, I assume that that one's overall wellbeing level at any world where one never exists at all is zero. See parts 2.2.1–2.2.3. This assumption is, though controversial, one we should be willing to make for purposes here. For something like it plausibly will be part of the case for the claim that a2 makes things better for Bob than a1 does and that a1 accordingly imposes a *loss* on Bob. That case is at least cogent enough that we would not want to rest our own account of what ought to be done on the assumption that it can be safely rejected. Nor do we want to beg the question against Inclusion.

<sup>58</sup>These natural numbers, with the exception of zero, need have their ordinal value only. By *overall wellbeing levels*, I mean *lifetime* wellbeing levels, taking into account all the benefits and all the burdens that ever come to the specified person at the specified world. More generally, *wellbeing* for purposes here means whatever it is that makes life so precious to the one who lives – not just subjectively precious, but truly of value, to that person. It seems to me that a capability approach to value will be important in defining wellbeing. See for example Sen (1992). But we do not want to overlook that the subjective component is important, too – that capability combined with joy is better for a person than capability alone (even if we can't figure out a way to say that joy itself is a part of capability). For purposes here, however, I set aside whether this way of looking at wellbeing is plausible. For a more recent discussion, see Feldman (2009).

not a2. After all, a2 creates more wellbeing for Bob than a1 does, since Bob by hypothesis has an existence “well worth having” at w2 and has nothing at all – no benefits; no burdens – at w1.<sup>59</sup>

Our intuitions about what ought to be done in the Basic Case seem clear: in some way *Ann must come first*. a1 is obligatory, and a2 is wrong. We ought to sacrifice the interests – such as they are – of the *possible* person Bob since doing so is *necessary* to protect the interests of the *existing* Ann. This is not to claim that we will not, by dint of argument at the end of the day, find ourselves forced to abandon intuition and adopt another view. This is just to claim that, starting out, we intuitively think that a1 is obligatory and a2 is wrong.

## 2.4 Exclusion Alpha

2.4.1 The *truly actualist* form of Exclusion – what I call *Exclusion Alpha* – provides that only those persons who do or will exist in what is the uniquely *actual* world matter morally.

We can state Exclusion Alpha as follows.

*Exclusion Alpha:*

Losses incurred at the *actual world* *w* or at any other accessible world *w'* by any person *p* who does not and will not exist at *w* – including the loss incurred by *p* when agents fail to bring *p* into existence – have *no moral significance* for purposes of determining the permissibility of any act *a* performed at *w* that imposes those losses or any alternative act *a'* performed at any world *w'* that avoids those losses.

A loss incurred at the *actual world* *w* or at any other accessible world *w'* by any person *p* who does or will exist at *w* is morally significant for purposes of determining the permissibility of an act *a* performed at *w* that imposes that loss and any alternative act *a'* performed at any other world *w'* that avoids that loss.

---

<sup>59</sup>Again, I assume that the wellbeing level a person has a world where that person never exists at all is just zero. See part 2.2.1–2.2.3 above.

Nonidentity worries can arise at this point. The supposition that the act a2 will eventually lead to the conception and coming into existence of any particular person *Bob* (as opposed to any one of many, many nonidentical others) will hold only on the rarest of occasions. More typically, we can be confident only that an act will lead to the existence of *some person or another*. To insure that the discussion that follows applies in both sorts of situations, we can understand the term “Bob” itself to function as a constant, not a name. Thus, we are saying that a2 will produce a certain sort of outcome at the world w2 where a2 is performed. That outcome will include some person or another – call that person “Bob” – who will have an existence well worth having; and that same person will not exist at all if a1 is performed instead. At the end of the argument, we can then discharge the constant “Bob” in accordance with the normal rules for existential derivation without any adverse impact on the conclusion we mean to draw – whether that conclusion is that a1 is obligatory in view of its effect on Ann, or a1 and a2 are both permissible in view of their effects on Ann and some other person, or something else entirely.

According to Exclusion Alpha, the effects of our acts on actual persons and actual persons alone are to be taken into account in determining what we do. Their losses, and their losses alone, have moral significance for purposes of evaluating the acts that impose those losses and the alternative acts that avoid those losses.

2.4.2 Exclusion Alpha won't tell us anything about the Basic Case until we add one additional fact. We need to say which world is *actual*.

Let's suppose that w1 is the uniquely *actual* world – and that a1, accordingly, is the act *actually* performed and that Ann *actually* exists while Bob remains *merely possible* relative to the actual world.

Let's then begin by looking at the loss Bob incurs at w1. Since w1 is actual and Bob never exists at w1, Exclusion Alpha implies that the loss Bob incurs at w1 has *no* moral significance for purposes of evaluating a1. Nor does that loss have any moral significance for purposes of evaluating a2 *even though Bob exists at w2*. Bob's loss at w2 is thus, according to Exclusion Alpha, devoid of moral significance across the board.

In contrast, since Ann exists at the actual w1, Exclusion Alpha implies that the loss she incurs at w2 has *full moral significance*. It *bears on* the permissibility of *both* a1 and a2. It *counts against* a2. In contrast, a1 altogether avoids that loss on Ann's behalf – not just in a vacuous way, by leaving Ann out of existence altogether, but rather in a way that makes things *better for Ann*. The otherwise plausible permissibility theory – specifically, OPPP1 – thus directs that a2 is wrong. Moreover, since a1 imposes no morally significant loss on anyone at all – on any existing, future or merely possible person; no loss at all on Ann, and no morally significant loss on Bob – OPPP2 instructs that a1 is permissible.<sup>60</sup> Standard deontic axioms tell us more: a1 is not just permissible, but *obligatory*, given that a2 itself is wrong.

And that result is a plus for Exclusion Alpha. At the very least, Exclusion Alpha comports very nicely with our original intuitions about the Basic Case.

2.4.3 But a slight change in the hypothetical turns this moral evaluation on its head. Before we added the fact that w1 is actual. Let's now instead add the fact that w2 is actual. It is a2 that is actually performed, and it is Ann *and Bob* who actually exist, not just Ann. Exclusion Alpha now instructs that the losses incurred by both Ann and Bob – Bob at w1 and Ann at w2 – bear on the permissibility of both a1 and a2. The case becomes, under Exclusion Alpha, a *wash* – a case involving acts that generate losses that are equally great and equally deep *and* have full moral significance, a case where it just does not matter what the agents choose to do. OPPP3 accordingly implies that a2 is *permissible*. And so is a1.

2.4.4 Exclusion Alpha thus conforms to the clear intuitions we have about the Basic Case only about half-way. When, by hypothesis, w1 is actual and a1 is actually performed, we obtain the intuitive result that a1 is permissible, in fact, obligatory, and a2 is wrong. But when we change the hypothetical and instead suppose that w2 is actual and a2 is actually performed, we suddenly obtain the result that a1 and a2 are *both permissible*.

But *does* the new hypothetical really trigger a new intuition? Does a2 – and it's the same old act, just situated within a new case; it's the same old act, now made

<sup>60</sup>We could, but need not, think of such deontic axioms as part of the otherwise plausible permissibility theory. See part 2.2.4 above.

actual when before we stipulated that it was left *unperformed* at the actual world – now seem *permissible*? We might just adore Bob, if he exists and we, too, exist. But our intuition as to permissibility is surely exactly what it was in the original case: a2 is wrong.

Exclusion Alpha has already committed itself to the view that if w1 is actual and a1 is actually performed – that is, if a2 is left *unperformed* – then a1 is obligatory and a2 is wrong. It is hard to understand how the mere *performance* of a2, which we just said was wrong if *unperformed*, could make a2 *right*.

The more plausible view is surely that the permissibility of an act does not depend on whether that act is actually performed. Put another way, Exclusion Alpha violates what Wlodek Rabinowicz calls the *Principle of Normative Invariance*.<sup>61</sup>

2.4.5 What we are seeing in Exclusion Alpha can be described as a *conceptual instability*, a situation in which the evaluation undergoes a seismic shift as a result of something that – intuitively – has more to do with history or metaphysics (was the dastardly act in fact performed or not?) than morality.

Now, the supposition that w1 is actual and that a1 is actually performed, and the alternate supposition that w2 is actual and that a2 is actually performed, cannot both be correct in any one case. There is only one *actual* world, and the performance of a1 precludes the performance of a2. Thus we are really talking about two distinct cases – the Basic Case, where w1 is actual, and an alternate version of that case, where w2 is actual instead. So the conceptual instability that we see does not itself suggest any logical inconsistency.

But that does not mean that the problem is not a very serious one. We have already said that Exclusion Alpha conforms only about half-way to the intuitions we have regarding the Basic Case. And now we see that Exclusion Alpha reads tea

---

<sup>61</sup> According to that principle, the permissibility of an act does not depend on whether that act is performed. The principle is attributed to Rabinowicz by Eric Carlson (1995), p. 100. The difficulty with theories that fail to conform to the Principle of Normative Invariance is, according to Carlson, that they are not “action-guiding.” Carlson (1995), p. 101. It is not beyond conception that a plausible permissibility theory may sometimes for any number of reasons fail to guide action. A theory can, for example, be plausible without being *complete*. Moreover, actual value forms of consequentialism may be less adept at guiding action than are their expected value analogs. But it is probably a mistake to reject actual value theories across the board just on the ground that they often fail to guide action; after all, expected value approaches often themselves fail to point us precisely in the direction of what we ought to do or permissibly may do. In my own view, then, the Principle of Normative Invariance goes still deeper: it underlines that whether an act is performed or not is not itself a morally salient property. A *wrong* but unperformed act seemingly *can't* be made *permissible* by nothing more than the agent's performance of that act. Many facts are not pertinent to moral permissibility; the work of constructing a moral theory will be successful or not depending on our ability to attend to the facts that are pertinent and, just as important, *to disregard the rest*. Before we can accept, in other words, that performance is pertinent to permissibility, we need an argument – or at least a handful of cases that compel us to appreciate the connection on an intuitive basis. Certainly, however, none of the cases we shall be examining here make that connection. The cases under consideration here, if anything, suggest just the reverse: that what is going on in these cases has nothing to do with whether the act is actually performed or not. For critical discussion of the Principle of Normative Invariance, see Frances Howard-Snyder (2008), pp. 1–15.

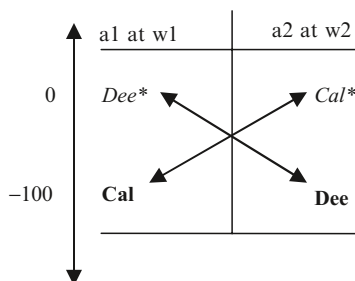


leaves that we ourselves cannot connect with what we are trying to explain. It thinks performance is important, and we can't see how that is so.

2.4.6 There's still another problem with Exclusion Alpha. It leads us to the result that either act, if *actually* performed, is permissible. But the whole point of Exclusion Alpha was to *put Ann first* – to assert, that is, that it *does* matter in this kind of case whether we avoid the loss on Ann's behalf or not, that it's not “just as good” to avoid the loss on behalf of Bob. This means that Exclusion Alpha isn't doing the work that it was meant to do. That in itself does not mean that Exclusion Alpha is *false*. Perhaps it was our initial intuitions about the Basic Case that were defective. But before we decide it's the intuitions that are defective and not Exclusion Alpha, we should see whether there is another way of making sense of those intuitions.

## 2.5 Double Wrongful Life

2.5.1 Other cases challenge Exclusion Alpha as well.<sup>62</sup> In *Double Wrongful Life*, the agent's choices are limited to bringing one or the other of two *wrongful lives* – lives, that is, that are *less* than worth living – into existence (Case 2.2).<sup>63</sup>



Case 2.2 Double Wrongful Life

Let's first suppose that  $w_1$  is the uniquely actual world and that  $a_1$  is actually performed. Cal thus actually exists and is accorded a wrongful life – a wellbeing level of  $-100$ . According to Exclusion Alpha, since Cal alone exists at  $w_1$ , the loss Cal incurs at  $w_1$  alone bears on the permissibility of  $a_1$  and  $a_2$ . The loss Dee incurs at  $w_2$  is to be screened out of the analysis altogether.  $a_2$  thus avoids a morally significant loss on behalf of Cal and makes things better for Cal – all without imposing a morally significant loss on anyone. On those facts, OPPP1 implies that  $a_1$  is wrong. OPPP2, moreover, instructs that  $a_2$  is permissible. Standard deontic axioms tell us still more: that  $a_2$  is obligatory.

<sup>62</sup> The objections against Exclusion, in both its Alpha and Beta forms, that I explore here have been discussed elsewhere. See Arrhenius (2009) and Caspar Hare (2007). See also Parsons (2002), pp. 135–47 and, generally, Parfit (1987), part four.

<sup>63</sup> Case 2 is based on Hare's Jack and Jane case, which he describes as “negatively symmetrical.” Hare (2007), p. 504.

2.5.2 But our intuitions surely go in the other direction. The very symmetry of the case tells us that *a1* and *a2* should be similarly evaluated. Either both acts are wrong or they are both right.<sup>64</sup> But it seems that moral law must always give agents at least one permissible alternative. So intuition alone leads us to think that *a1* and *a2* are *both* permissible.

2.5.3 There's an even deeper problem. Let's now change hypotheticals. Let's now suppose that it is *a2* that is actually performed rather than *a1*. Exclusion Alpha now, by a parallel analysis, generates the result that *a2* is wrong and *a1* obligatory.

But we have already said that, in the original version of the case, the version in which *w1* is actual, that *a1* is wrong and *a2* obligatory. And that means that, whether the agent actually performs *a1* or actually performs *a2*, what the agent does is wrong *and* what the agent could have done instead is permissible.

2.5.4 If this looks like an inconsistency in Exclusion Alpha, it isn't. For we are really dealing with two distinct cases, one that includes the supposition that *w1* is actual, and one that includes the supposition that *w2* is actual.

However, we do face here exactly the same sort of conceptual instability we saw in the Basic Case. If *a1* is actually performed, *a2* looks permissible. But the second we shift cases, and suppose instead that it is *a2* that is actually performed rather than *a1*, the evaluation of *a2* also shifts.

2.5.5 One last point. It's true that both variations on Double Wrongful Life leave the agent with a permissible alternative. That means that Exclusion Alpha avoids the classic moral dilemma. It complies, that is, with the idea that moral law always gives agents at least one permissible alternative. The problem is they can't *perform* that alternative and have it *stay* permissible. In the end, whatever they actually do is sure to be wrong.

## 2.6 Addition Plus

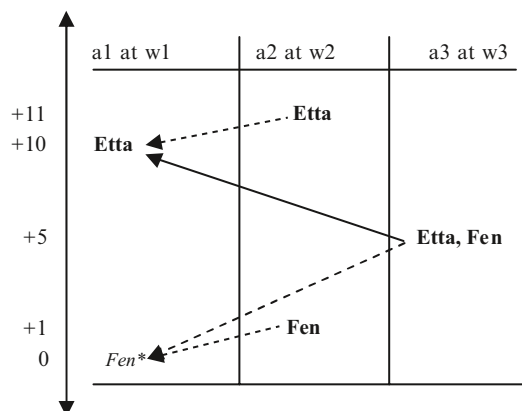
2.6.1 Still another problem case for Exclusion Alpha is *Addition Plus*.<sup>65</sup> Addition Plus is like the Mere Addition Paradox in that an additional person whose life is worth living is added to the mix.<sup>66</sup> The "Plus" is that adding the additional person here makes *someone else better off*.

---

<sup>64</sup>The differences between *a1* and *a2* constitute *merely reversing changes*. See generally Peter Vallentyne (2000), pp. 1–19.

<sup>65</sup>I have elsewhere argued that this sort of case shows that such overly simplistic sufficient conditions on impermissibility – however person-based in nature they may be considered to be – must be rejected. See Roberts (1998), pp. 63–64.

<sup>66</sup>Addition Plus raises many of the same issues as one of the problem cases that Parfit and Singer discussed in connection with a proposal explored by Singer. That proposal was an attempt to articulate an alternative to the Totalist treatment of additional person cases generally. See Singer (1976), p. 97, and Parfit (1976), p. 105. See also Parfit (1987), pp. 419–41. More recently, Singer articulated and rejected what he called the *Prior Existence View*. See Singer (1999), pp. 102 ff. For discussion of the Prior Existence View, see part 2.11 below.



Case 2.3 Addition Plus

Let's begin with the supposition that  $w_1$  is the actual world and  $a_1$  is actually performed. Then, according to Exclusion Alpha, since Etta alone exists at  $w_1$  and  $w_1$  is actual, Etta's losses alone have moral significance for purposes of evaluating each of the agent's alternatives – not just  $a_1$  but  $a_2$  and  $a_3$  as well. OPPP1 implies, accordingly, that  $a_1$  and  $a_3$  are both wrong; and OPPP2 that  $a_2$  is permissible. Standard deontic axioms tell us, as well, that  $a_2$  is obligatory. After all,  $a_2$  avoids not one but two morally significant losses on Etta's behalf, and makes things better for her than either  $a_1$  or  $a_3$  does. It's true that Fen incurs a loss at  $w_2$ . That he does is established not by comparing how well Fen fares in  $w_2$  against how well he fares in  $w_1$ , but rather by comparing how well Fen fares in  $w_2$  against how well he fares in  $w_3$ . The former comparison demonstrates no loss at all; the latter, a very great and very deep loss. But since Fen never exists at  $w_1$ , Exclusion Alpha implies that that very great and very deep loss is *devoid of moral significance*.<sup>67</sup>

2.6.2 But the results that  $a_1$  is wrong and  $a_2$  obligatory intuitively seem false. We cannot really think that it is wrong for agents to refrain from bringing a new person into an avoidably bad existence all in order to secure some slight gain for one person. Since it is part of the case that agents have the option, at  $w_3$ , of doing so much more for Fen, we cannot really think that it is obligatory, or even permissible, for agents to bring Fen into existence *and do so much less for him*, all to secure some slight benefit for Etta. That it would have created additional wellbeing for my

<sup>67</sup> Now, we may well consider wellbeing more appropriately distributed between Etta and Fen under  $a_3$  than under  $a_2$ . But the tradeoff principles an otherwise plausible permissibility theory can be expected to include will never require the agent to impose a morally *significant* loss on one person – Etta in  $w_2$  – so that the agent can avoid imposing a morally *insignificant* loss on someone else – Fen in  $w_1$ .

child to have her own private live-organ donor, or slave, does not make it wrong for me *not* to have brought an additional person into existence *and to have avoidably treated that person badly*.<sup>68</sup> There are limits to what I may do even when the person I am doing more for is *my own child*.

2.6.3 Of course, some philosophers – inspired, perhaps, by a total, or pluralistic, form of aggregative consequentialism, and notwithstanding the point that both Totalism and Pluralism are quite at odds with Exclusion Alpha<sup>69</sup> – may want to press the argument that a1 *is* wrong. I don't think a1 *is* wrong. But if it is, we should note that *even the Totalist* agrees that any plausible view must *at least* take Fen's loss at w2 into account as it makes its way toward a determination of a1's permissibility. Fen's loss at w2 at least *bears on* the permissibility of a1. That modest point is, however, one that Exclusion Alpha must reject in the case where w1 is actual.

2.6.4 Many of us have the following intuitions. (A) Etta is not compelled to suffer just so that the merely possible Fen can come into existence and be treated well, as Fen is under a3. But (B) *if* Fen is brought into existence at all, the fact that a3 is a perfectly viable alternative way of bringing Fen into existence means that a2 itself is far too flawed to establish that a1 is wrong. Affirming both (A) and (B) means that a1 is (at least) permissible. But it also means that the loss Fen incurs at w2 *bears on* our evaluation of the permissibility of a1 – and that that is so, notwithstanding the fact that Fen never actually exists at w1. By counting *against* a2, Fen's loss at w2 in a roundabout way thus counts *in favor of* a1.

Fen's loss at w2 thus *seems*, at least, enough to make the otherwise obligatory but nonactual a2 wrong and the otherwise wrong but actual a1 permissible. But if that is correct, then we shall have no choice but to reject Exclusion Alpha, which insists that Fen's loss at w2 has *no bearing at all* on the permissibility of a1.

2.6.5 Nils Holtug suggests that views like Exclusion Alpha, by virtue of the claim that persons who are merely possible relative to the actual world *never* matter morally, have more in common with "racism, sexism and speciesism" than we would like.<sup>70</sup> Holtug's language seems a bit over the top until we come to understand

---

<sup>68</sup> The nonidentity problem arises here. It may seem to undermine the assertion that a3 is better for Fen than a2 is. Can we really make it *part of* Addition Plus that agents have the genuine alternative in a3 of creating additional wellbeing *for Fen*, the very same, identical person who is made to suffer under a2? The simple answer here is that it's part of the hypothetical that agents in fact do have that very option. The more interesting answer is that the very type of nonidentity problem that might be used to challenge the hypothetical is in fact based on a fallacy. See part 2.2.5 above and Appendix B.

<sup>69</sup> See note 36 above (description of Pluralism).

<sup>70</sup> Holtug (1999), p. 26. It is not necessarily that we have Kantian objections against, for example, slavery. It might just as well be that we are consequentialists of one sort or another and think that the option of bringing someone into such an impoverished existence – with so little justification; just 1 unit of wellbeing for the already very well-off Etta – when we had the option of *doing so much more* for that same person cannot establish that the alternative of leaving that person out of existence altogether is wrong.

just what a fully-charged commitment to existing and future persons obligates us to do to the merely possible.

The difficulty is not at all that Exclusion Alpha on occasion obligates us to *leave the merely possible out of existence altogether* – when, for example, doing so makes things better for some existing or future person without making things worse for anyone at all. The difficulty is that Exclusion Alpha on occasion also makes it our obligation to bring the merely possible into existence and *unjustifiably* create less wellbeing for those same (“identical”) persons when we could have created more. But surely that is not an obligation we in fact have.<sup>71</sup>

2.6.6 Caspar Hare describes still another case that shows that Exclusion Alpha struggles with nonactual acts whose adverse effects are limited to nonactual persons. Exclusion Alpha – what Hare calls *Strong Moral Actualism* – seems to instruct that the “genocidal adventures of nonactual dictators” are permissible, or even obligatory, in cases where the victims of those adventures are themselves nonactual.<sup>72</sup> Exclusion Alpha leaves the otherwise plausible permissibility theory with no grounds for declaring such “genocidal adventures” wrong. But we think that they are.

2.6.7 That we intuitively consider wrong the terrible things Darth Vader did to *his own son* might seem reason enough to reject Exclusion Alpha. In defense of Exclusion Alpha, however, it might be argued that we should not be too concerned with how the evaluation of *nonactual* acts comes out. It may seem that we should not be particularly troubled when the theory fails to produce satisfying results for acts *whose victims are merely hypothetical*. It may seem that what is important is that the theory produce satisfying results for acts *whose victims are actual*. And, indeed, had the “genocidal adventures of nonactual dictators” in fact taken place, Exclusion Alpha, in combination with any otherwise plausible permissibility theory, would no doubt have judged those adventures wrong since their victims would then have been actual.<sup>73</sup>

This defense of Exclusion Alpha can be put another way. Exclusion Alpha comes with a caveat. If we *had* actually done as we ought – if we had actually brought those merely possible persons into existence *and* treated them badly – then those persons would not have been *merely possible* but rather *actual*. If, in other words, we revise the case in such a way that the act *actually* performed, and the class of persons who do or will exist at the uniquely *actual* world shifts – if, in other words, we *had* done as we ought – then Exclusion Alpha would have declared what we have done wrong.

The caveat thus insures that the “genocidal adventures of nonactual dictators” *would have been wrong* had those dictators and their bad acts been actual rather than not.

Similarly, it’s true that, in Addition Plus, where a1 is actually performed and Etta and only Etta accordingly matters, a2 is obligatory. But the caveat insures that, had a2 actually been performed in lieu of a1, a2 *would have been wrong*.

---

<sup>71</sup> Holtug (1999), p. 26.

<sup>72</sup> Hare (2007), 503. But see Parsons (2002), pp. 135–47.

<sup>73</sup> See generally Parsons (2002), p. 145.

2.6.8 But it is not clear whether this caveat rescues Exclusion Alpha from its own excesses or reveals just how deeply flawed Exclusion Alpha really is. The difficulty is that Exclusion Alpha *cannot be contained to questionable evaluations of nonactual acts and the nonactual losses they impose*. The difficulty extends to how *actual* acts, imposing *actual* losses, are evaluated as well.

Addition Plus makes this point. Part of the reason that the *actual* a1 intuitively seems permissible in Addition Plus, notwithstanding the *actual* loss it imposes on Etta, is what the *nonactual* a2 does to the nonactual Fen. When Exclusion Alpha instructs us to exclude Fen's interests from our calculation of the permissibility of a1, it forces us to say that the additional wellbeing a2 creates for Etta is *sufficient* to show that a1 is wrong. Thus a1, according to Exclusion Alpha, is not simply wrong *if* performed. It's both performed *and* wrong. The upshot is that the deficiency we see in Exclusion Alpha cannot be contained to nonactual acts and the nonactual losses they impose.

Thus, the caveat shows that, had another world and another act been actual, we would have been able to obtain, under Exclusion Alpha, a more plausible permissibility result. That's all well and good. The problem is that it doesn't in the least change the fact that we have already drawn from Exclusion Alpha a permissibility result that is not plausible at all: that the *actual* a1 is wrong.

2.6.9 The conceptual instabilities that we earlier noted in connection with Exclusion Alpha are repeated in Addition Plus. Suppose that w1 is the actual world and that it is a1 that is actually performed. As a simple conceptual matter, it is hard to understand how our performing the nonactual act a2 that is itself, if unperformed, permissible or even obligatory – *as we ought if it's obligatory* – can make that act *wrong*. The better view seems to be that, if mere performance makes the act wrong, surely it was never *genuinely obligatory* to begin with. But if we take that view, we are forced to reject Exclusion Alpha.

2.6.10 There is a related problem. Addition Plus, like Double Wrongful Life, underlines that Exclusion Alpha cannot be counted on to give us a *way out of wrongdoing*. *Not* bringing the merely possible Fen into existence and treating him badly is wrong in virtue of its being bad for Etta. But *bringing* Fen into existence and treating him badly – had we actually done that instead – would have been just as wrong. We again have a case in which the agent can *do no right*. This is not technically a moral dilemma, but it doesn't really adhere to the spirit of "ought implies can," either. Whatever the agent does, he or she has a morally permissible alternative. The problem is that the agent can't *perform* that alternative and have it *stay* permissible.

2.6.11 Why is Exclusion Alpha failing in these cases? The obvious point is that Exclusion Alpha is *too exclusive*: it is screening out morally significant losses. It thus fails to recognize in Double Wrongful Life that the loss Dee incurs at w2 is what makes a1 *permissible*; her loss counts, not just *against* a2 in the case where a2 is actually performed, but, in a roundabout way, *in favor of* a1. Similarly, even if Cal never actually exists, the loss he incurs at w1 is what makes a2 *permissible*. Addition Plus makes the same point. The loss Fen incurs at w2 clearly at least *bears on* our evaluation of both a1 and a2 even if a1 is the act actually performed and Fen remains merely possible. It is that loss that makes a2 clearly wrong – and, in a roundabout way, a1 plausibly *right*.

2.6.12 But the more subtle point is that Exclusion Alpha is *not exclusive enough*. At certain points, it allows the merely possible to exert more of a moral pull on what the agent does than they ought. According to Exclusion Alpha, all that has to happen, for the merely possible to matter morally, is for the supposition regarding which act is actually performed to be modified. At first glance, that feature seems desirable – or at least innocuous. If the *merely possible* are *made actual* by changing the facts of the case, then surely they will then come to matter morally. If nothing else, surely we can agree that *all actual people always matter morally*.

I believe, however, that this is one of those many cases in which a superficially obvious truth is false. It is a mistake, that is, to think that actual persons matter morally in some *necessary* way – to think, that is, that *all* their losses have moral significance. The Basic Case makes that point. Where a1 is actually performed, Exclusion Alpha functions just as it should. It goes awry when we shift to the supposition that a2 is actually performed. Then Exclusion Alpha abruptly assigns moral significance to *all* of the losses incurred by Bob *wherever* he incurs them, *including* at w1 where he never exists at all. But the fact that Bob is actual hardly suffices to make a2 permissible. We can *just adore* the actual Bob but still consider what has been done to Ann wrong and still insist that what we ought to have done was leave Bob out of existence altogether.

Perhaps we are pleased that we ourselves happened to have been conceived against all odds and then born.<sup>74</sup> I certainly am. But that is not enough to show that we *matter morally in some necessary way* – that all of our losses, wherever we happen to incur them, have full moral significance or indeed any moral significance at all. That is not enough to show that the loss we would have incurred had we never existed at all *counts against* whatever it was our parents might have done in lieu of bringing *us* into existence.

The same point holds for the Basic Case. Bob will exist if a2 is performed. But whether he exists or not, the loss he incurs at w1, where he never exists, has no moral significance. The puzzle, of course, is how to say that *without* implying that it is permissible or even obligatory, in Addition Plus, to bring Fen into existence and then treat him badly.

2.6.13 The cases thus suggest that Exclusion Alpha is both too exclusive *and* not exclusive enough. I think that means that it's not going to work either to loosen *or* tighten Exclusion Alpha. It means that we need a mechanism for doing both at the same time. But is that a mechanism either Exclusion Beta or Inclusion can provide?

---

<sup>74</sup>“Imagine a speck of dust next to a planet a billion times the size of the earth. The speck of dust represents the odds in favor of your being born, the huge planet would be the odds against it. So stop sweating the small stuff. Don't be like the ingrate who got a castle as a present and worried about the mildew in the bathroom. Stop looking the gift horse in the mouth – remember that you are a Black Swan.” Taleb (2007), p. 298.



## 2.7 Exclusion Beta

2.7.1 For Exclusion Beta, the critical fact is not who *is* actual and how an act and its alternatives affect those persons, but rather who *would be* “actual” if the act we aim to evaluate had been performed. Under Exclusion Beta, just how a *nonactual* act affects *nonactual* persons becomes just as important as how an *actual* act affects *actual* persons. The only restriction is that those nonactual persons must exist, at least eventually, *under that nonactual act* for their losses to have moral significance for purposes of evaluating that act.

Exclusion Beta can be put as follows.

*Exclusion Beta:*

Losses incurred at any world  $w$  or at any other accessible world  $w'$  by any person  $p$  who does not and will not exist at  $w$  – including the loss incurred by  $p$  when agents fail at  $w$  to bring  $p$  into an existence worth having – have *no moral significance* for purposes of determining the permissibility of any act  $a$  performed at  $w$ .

Losses incurred at any world  $w$  or at any other accessible world  $w'$  by a person  $p$  who does or will exist at  $w$  are morally significant for purposes of determining the permissibility of an act  $a$  performed at  $w$ .

A person thus matters morally for purposes of evaluating a particular act *if and only if* that person does or will exist at the world at which that act is performed. This means a person who counts as merely possible relative to the *actual* world may still matter morally. That person will matter morally if he or she does or will exist at the world at which the act under scrutiny is performed.

2.7.2 Exclusion Beta starts out well enough in the Basic Case. At least: Exclusion Beta evaluates  $a_1$  just as Exclusion Alpha does when combined with the supposition that  $w_1$  is the actual world. (And Exclusion Beta doesn't need the supposition.) Since  $a_1$  is the act under scrutiny and Ann and Ann alone exists if  $a_1$  is performed, Exclusion Beta implies that the loss Ann incurs at  $w_2$  is morally significant for purposes of evaluating  $a_1$  but the loss Bob incurs at  $w_1$  is not. The analysis under the otherwise plausible permissibility theory proceeds just as it did under Exclusion Alpha. We obtain the result that  $a_1$  is obligatory – a result that perfectly comports with our intuitions about the Basic Case.

2.7.3 But when we turn to  $a_2$ , Exclusion Beta quickly runs into trouble. Both Ann and Bob exist if  $a_2$  is performed. Exclusion Beta thus implies that the losses incurred by Bob at  $w_1$  as well as by Ann at  $w_2$  have full moral significance.

The analysis now proceeds just as it did for Exclusion Alpha when combined with the supposition that it is  $a_2$  that is actually performed. The upshot is that Exclusion Beta, just like Exclusion Alpha, considers the case a *wash*: a case involving acts that generate losses that are equally great and equally deep *and* have full moral significance. The otherwise plausible permissibility theory will direct that  $a_2$  is permissible.

That result seems clearly counterintuitive. Our intuition is that Ann must come first. Exclusion Beta now instructs that Ann and Bob both matter morally – more

precisely, that all the losses each incurs, whether at w1 or at w2, have full moral significance now that we ourselves have shifted our own focus from a1 to a2.

2.7.4 And there is a second problem. It is not just that it is *counterintuitive* to say that a2 is permissible. It is also that we said just before that a1 is *obligatory*. These are inconsistent results. At least, they flout deontic axioms we may be loathe to reject. To say that an act is *obligatory* is to say that agents must perform *that act and no other*. But to say an alternative act is permissible is to say that agents *may perform that alternative act*. Accordingly, if a1 is *obligatory*, a2 *cannot* be *permissible*.

We said before that Exclusion Alpha was conceptually unstable in a certain way. We evaluated a1 one way given the supposition that w1 was actual and another way given the supposition that w2 was actual. More precisely, we evaluated a1 as obligatory in the version of the case that included the supposition that w1 was actual, and as permissible in the *alternate version* of that case that included the supposition that w2 was actual. But the situation we see in Exclusion Beta is, if anything, a bit worse. We now see a shift in evaluation taking place in the *very same case*. So here we face not “just” a conceptual instability but – taking deontic axioms into account – an out-and-out inconsistency.

The roots of the inconsistency are perfectly clear. Exclusion Beta claims that the class of persons who matter morally can shift, depending on which act happens to be under scrutiny. But we also think that how we evaluate one act constrains how we evaluate another. Deontic axioms – not to mention intuition – forbid our saying in one and the same case both that one act is obligatory and that an alternative act is permissible. Thus, in the Basic Case, Exclusion Beta in combination with the otherwise plausible permissibility theory judges a1 *obligatory* relative to the collection of persons who happen to exist under that act – that is, Ann. But Exclusion Beta *in the very same case* judges an alternative act *permissible* relative to the *distinct* collection of persons who happen to exist under that *distinct* act – that is, Ann and Bob. We thus find ourselves saying that a1 is *obligatory* and a2 *permissible*.

2.7.5 Exclusion Beta also stumbles badly in Double Wrongful Life. Thus, when we evaluate a1, we obtain from Exclusion Beta that the loss Cal incurs at w1 and that loss alone has moral significance for purposes of evaluating a1. It's true that Dee incurs a loss at w2. But, according to Exclusion Beta, Dee's loss has no moral significance since it is a1 we are evaluating and Dee never exists at w1. The otherwise plausible permissibility theory – specifically, OPPP1 – accordingly implies that a1 is wrong.

But that result is counterintuitive. By whatever measure, a2 is obviously just as dreadful as a1 is. There is nothing in the very unfortunate a2 that even begins to make us think that the very unfortunate a1 is wrong. We appreciate that a1 has a very bad effect. But it is rash to jump from that fact to the claim that a1 is wrong. We only make the mistake of doing that if we've failed to inspect a1's alternatives in any careful way. When we do, we decide that a1 is permissible.

2.7.6 Things get worse when we turn to evaluate a2. An exactly symmetrical analysis implies that a2 is wrong. But we've just said that a1 is wrong. Of course these results track the results we obtained under Exclusion Alpha. But we've now obtained the results that a1 and a2 are both wrong in the *very same case*, not two versions of a case that differ in the suppositions they make regarding which is the actual world. Exclusion Alpha meant that the agent always had a permissible option,

even if that option could not be both actually performed and remain permissible. Exclusion Beta leaves the agent with a classic moral dilemma – a case in which the agent has *no* permissible alternative.

But can that ever happen? A standard deontic axiom says it can't. Now, one might want to challenge that axiom. One would do that by presenting an extreme case in which we are strongly and irremediably pulled to say that both alternatives are wrong. But Double Wrongful Life is just not that kind of case. It's an odd case, but we are not pulled to say that *both* options are wrong. We may think both options are *horrible*, but that is different. That is not enough to make us think that agents have been left with *no permissible alternative* at all.

The discussion would go very differently, of course, if we created a completely new case by giving agents still a *third option*, a3, that would leave both Cal and Dee out of existence altogether (or bring them both into existence and give them lives worth living, or etc.). But that's not *this* case – nor would that new case create a moral dilemma. We would just say that a1 and a2 are wrong, and that a3 is permissible and, indeed, obligatory.

2.7.7 Now, Exclusion Beta does have what seem to be some clear advantages over Exclusion Alpha. It seems a plus for Exclusion Beta that it, in combination with the otherwise plausible permissibility theory, can declare nonactual acts whose bad consequences are limited to nonactual persons wrong. Exclusion Beta recognizes that the losses incurred by persons who are *merely possible relative to the actual world* have moral significance for purposes of evaluating any nonactual act under which those persons do or will exist. Exclusion Beta, unlike Exclusion Alpha, can thus declare the “genocidal adventures” of Caspar Hare’s “nonactual dictators” wrong.

2.7.8 This feature may make it look like Exclusion Beta does a better job than Exclusion Alpha with nonactual acts and their nonactual victims. Of course, the bare fact that we've already seen very deep conceptual difficulties in Exclusion Beta should make us skeptical that that is so.

What we should like and not like about Exclusion Beta becomes very clear in connection with Addition Plus. Even if w1 is the actual world and Fen never actually exists, Exclusion Beta recognizes that the loss Fen incurs at w2 – the loss Fen incurs not compared to how well he does in w1 (which is not well) but rather compared to how well he does in w3 (which is well) – is morally significant for purposes of evaluating a2. Etta also exists at w2. So the losses she incurs, at both w1 and w3, bear on the permissibility of a2 as well. Clearly, however, the loss Fen incurs at w2 is much *deeper* than any loss Etta incurs at w3. And that is so, whether we are talking about the loss she incurs at w3 relative to w1 or the loss she incurs at w3 relative to w2. The *volumes* of the various losses are pretty close. But, since Fen is left far worse off at w2 than Etta is at w3, the *depth* is very different. On these facts, the otherwise plausible permissibility theory may well deem a2 wrong, given that agents had the option of performing a3 in its place. The thought would be that it does not make the necessary tradeoff between Etta and Fen in an appropriate way.

Now, one might well quarrel with the underlying assumption that any permissibility theory that is both otherwise plausible and more complete will include the idea that agents ought first see to the needs of the least well off, then the next least well off, and

so on – will include, that is, a reference to Leximin. I like some of the variations on Leximin, but that's not a debate we need to have here. For purposes here, what is important is that the assumption that some such principle is part of the otherwise plausible permissibility theory is what shows Exclusion Beta off *to its best advantage*. It shows that Exclusion Beta can tell us what is wrong with a2. And, since we are likely, at an intuitive level, to think that a2 *is* wrong, we beg no question against Exclusion Beta if we assume that the otherwise plausible permissibility theory will contain a principle that deals with tradeoffs in a way that directs that a2 (given the option of a3) is wrong.

2.7.9 So far, so good. But so far *we have just evaluated* a2. When we turn to a1, Exclusion Beta does not do as well. It creates the same obligations to do the same avoidable and unjustifiably bad things to the merely possible that Exclusion Alpha does in the case where w1 is actual. According to Exclusion Beta, Etta's losses *alone* matter morally for purposes of evaluating a1 since she alone exists at w1. That is: the losses Etta incurs at w1 and w3, and those losses alone, have moral significance for purposes of evaluating a1. Moreover, a2 makes things better for Etta than a1 does. And it does so without – according to Exclusion Beta – imposing any morally significant loss on anyone else at all. Thus Fen incurs a loss at w2 (compared to w3). But that loss has *no moral significance* for purposes of evaluating a1, according to Exclusion Beta, since Fen never exists at w1. Given these moral data, OPPP1 instructs that a1 is wrong. But that result seems false.

So Exclusion Beta takes us to the same counterintuitive result that Exclusion Alpha does, given the supposition that w1, not w2, is actual. Exclusion Beta errs, just as Exclusion Alpha does, when it blinds itself to the loss that Fen incurs at w2 for purposes of evaluating a1.

2.7.10 Now, as noted earlier in connection with Exclusion Alpha, one might always argue that intuition has gone awry and that a1 *is* wrong. I don't think a1 is wrong. But even if it is, we should at least agree that Fen's loss at w2 *bears on* the permissibility of a1. Even the Totalist, bound to conclude that a1 is wrong, agrees that Fen's loss at w2 *bears on* the evaluation of a1. But according to Exclusion Beta, it doesn't. And that's a problem.

2.7.11 Like Exclusion Alpha, Exclusion Beta is at the same time both *too exclusive* and *not exclusive enough*. In both Double Wrongful Life and Addition Plus, Exclusion Beta is *too exclusive*. It screens out losses that are clearly morally significant. Dee's loss under a2 surely bears on the permissibility of a1. In a roundabout way, Dee's loss counts in favor of a1 by showing that a2 itself is too flawed to show that a1 is wrong. Cal's plight under a1 just as surely bears on the permissibility of a2 in the same roundabout way.

And in Addition Plus Fen's loss at w2 clearly counts, not just *against* a2, but also, in a roundabout way, *in favor of* a1. Fen's loss helps explain why we think a1 is permissible – or, at least, why we hesitate before we declare it wrong.

At the same time, in the Basic Case, Exclusion Beta is *too inclusive*. It does what it is supposed to do when we evaluate a1. But when we turn to a2, Bob's loss at w1 abruptly assumes *full moral significance* – forcing any otherwise plausible permissibility theory just as abruptly, having already declared a1 *obligatory*, now to say that a2 is *permissible*.

The better view is that the loss incurred by Bob at  $w_1$  is devoid of moral significance *regardless* of the act we happen to be evaluating. As noted before, the trick is to say that but *not* say, in Addition Plus, that it is permissible to bring Fen into existence and treat him badly.

## 2.8 Inclusion

2.8.1 Exclusion Alpha and Exclusion Beta both comport with our intuitions only about half-way. This means that the fact that Inclusion does not – as we shall see – *perfectly* comport with our intuitions does not give us a reason to prefer Exclusion. Moreover, Exclusion in both its forms faces serious conceptual difficulties – difficulties that Inclusion has the resources to avoid nicely. Thus, Inclusion avoids the mistake of linking an act's permissibility to its performance, and never even hints that an *obligatory* act can have *permissible* alternatives.

These realities have very likely led many theorists to consider Inclusion a serious alternative to Exclusion and – mistakenly, I think – a plausible view in its own right.

*Inclusion:*

For any person  $p$  who does or will exist at any world, any loss incurred at any world  $w$  by  $p$  – including the loss incurred by  $p$  when agents fail at  $w$  to bring  $p$  into an existence worth having – is morally significant for purposes of determining the permissibility of any act  $a$  performed at  $w$  that imposes that loss and any alternative act  $a'$  performed at any other accessible world  $w'$  that avoids that loss. That is so, regardless of whether  $p$  does or will exist at  $w$ .

According to Inclusion, *all* losses incurred by *all* persons at *all* worlds – including losses incurred at worlds where the persons who incur those losses never exist at all – matter morally for purposes of evaluating *all* acts, independent of who actually exists and independent of which act in fact happens to be under scrutiny. Inclusion thus casts a wide net on the morally significant loss.

2.8.2 Inclusion's intuitive and conceptual advantages are on display in Double Wrongful Life. Inclusion recognizes that Cal's loss at  $w_1$  is morally significant for purposes of determining the permissibility of  $a_2$ . Accordingly, the otherwise plausible permissibility theory that is combined with Inclusion does not blind itself to that loss in determining the permissibility of  $a_2$ . It can say, in particular, that the loss Cal incurs at  $w_1$  shows that  $a_2$  is really not any more dreadful than  $a_1$  is. The loss Cal incurs at  $w_1$ , by *counting against*  $a_1$ , in a roundabout way, thus counts *in favor of*  $a_2$ . The way is thus paved for the result that  $a_2$  is permissible. By the same logic,  $a_1$  is permissible as well. Those results seem clearly right.

2.8.3 In Addition Plus, Inclusion considers the loss Fen incurs at  $w_2$  – the loss established, of course, by comparing Fen's wellbeing level at  $w_2$  against Fen's wellbeing level at  $w_3$ , not  $w_1$  – morally significant for purposes of evaluating  $a_1$ . That is so, even if  $a_1$  is actually performed and Fen remains merely possible, and despite the fact that Fen would not exist if  $a_1$  were performed.

Now, we might quarrel about what a correct permissibility theory will do with that particular moral datum. Perhaps it will deem a1 wrong on the view that agents ought to maximize aggregate wellbeing, and consider a2 accordingly obligatory, despite the morally significant loss Fen incurs at w2. Or perhaps it will deem a1 wrong because it considers permissibility to be determined on the basis of *both* aggregate wellbeing *and* the distribution of wellbeing across the population. Taking that approach, we might say that a2 shows that a1 is wrong *and* that a3 shows that a2 is wrong but that a3 is itself permissible – that is, that there is nothing in a1 that shows that a3 is wrong. Or perhaps it will even find a way of saying that a1 is permissible or even obligatory, despite the morally significant loss Etta incurs at w1, on still another form of pluralism, a form that considers human flourishing a good, alongside aggregate wellbeing and equality.<sup>75</sup> What is most important for purposes here is that Inclusion *correctly* insists that the loss Fen incurs at w2 must be *taken into account* in evaluating a1; Fen's loss at w2 *bears on*, though does not on its own determine, whether a1 is permissible or not.

2.8.4 Inclusion avoids the conceptual difficulties that plague Exclusion Alpha and Exclusion Beta by making the class of losses deemed to have moral significance *fixed*. Double Wrongful Life and Addition Plus make that point. We encounter no surprises when we change suppositions regarding which act is actually performed and no inconsistencies when we shift from evaluating one act to evaluating another. Results remain stable, deontic axioms obeyed and inconsistencies avoided.

Moreover, by making the class of losses deemed to have moral significance *very large*, Inclusion avoids the intuitive difficulties Exclusion Alpha and Exclusion Beta create. Losses incurred by the merely possible – and those who would remain merely possible even had the act under scrutiny itself been performed – are understood to bear on permissibility. Double Wrongful Life and Addition Plus make that point as well. Thus Inclusion steadfastly recognizes that the loss Dee incurs at w2 bears on what is done at w1, the loss Cal incurs at w1 bears on what is done at w2 and the loss Fen incurs at w2 bears on what is done at w1. Intuitively, each of those losses does matter morally, not just in the sense that it counts against the act that imposes it, but also that it, in a roundabout way, counts in favor of the act that avoids it.

2.8.6 At the same time, the Basic Case shows that Inclusion is surely false. We intuitively think Ann must – in some way – come first. But that is not how Inclusion sees things. According to Inclusion, there is nothing to recommend a1 over a2. We can make people happy, we can make happy people. The otherwise plausible permissibility theory that comports with Inclusion will instruct that we may do as we please.

The Basic Case reveals something important about Inclusion: it puts the *merely possible in a wildly implausible competition with those who do or will exist*. The view that you have done something perfectly permissible when you refuse to rescue your own *real live dog* from a “terrible spat” for sake of the Calico Cat – or Thomasina or the Cat Who Went to Heaven or any other *merely possible cat* – seems just mistaken.

---

<sup>75</sup> See Temkin (1993).



In a way that can only be described as axiomatic, your actual dog *must* – in some way – come before any merely possible cat. We must find a way of saying that Ann comes before Bob even as we insist that Cal does not come before Dee (or vice versa) and that Etta does not come before Fen.

We must, in other words, find a way of saying (A) that Ann comes first *without* taking the position that (B) if  $w_1$  is actual in Addition Plus, and  $a_1$  is actually performed, then  $a_1$  is wrong. But we must also find a way of saying (A) without taking the position that (C) an act can be permissible, or even obligatory, if unperformed but wrong if performed. And finally we must find a way of saying (A) without taking the position that (D) an act is *obligatory* but that its alternative is – in the same case – somehow *permissible*. We must, in other words, avoid both Inclusion and Exclusion, in both of its forms.

## 2.9 Variabilism

2.9.1 A middle ground between Inclusion and Exclusion is Variabilism. The fact that Variabilism is at least plausible means that, having rejected the frying pan, we should not feel immediately compelled to leap into the fire.

Variabilism can be put as follows.

*Variabilism:*

A loss incurred at any world  $w$  by any person  $p$  has moral significance for purposes of determining the permissibility of any act  $a$  performed at  $w$  that imposes that loss and any alternative act  $a'$  performed at any accessible world  $w'$  that avoids that loss *if and only if*  $p$  does or will exist at  $w$ .

By implication, the loss incurred by  $p$  at  $w$  when agents fail to bring  $p$  into an existence worth having at  $w$  will have no moral significance for purposes of determining the permissibility of the act  $a$  performed at  $w$  that imposes that loss or any alternative act  $a'$  performed at  $w'$  that avoids that loss.

On this view, whether a loss bears on the permissibility of an act depends not on *who incurs that loss* but rather on *where that loss is incurred* in relation to the person who incurs it. For any particular person, some losses have moral significance for purposes of determining the permissibility of the act under scrutiny and its alternatives and some do not. A loss incurred by a person at a world is devoid of moral significance for purposes of evaluating the act that imposes that loss or any alternative to that act if that loss is incurred at a world where the person who incurs that loss never exists at all. Incurred, however, at a world where that person does or will exist, a loss will have full moral significance.

2.9.2 Variabilism functions in the Basic Case just as Exclusion Alpha does on the supposition that  $a_1$  is actually performed – and as Exclusion Beta does (given the inconsistency, perhaps we should say *tries* to do) when it is  $a_1$  that is under scrutiny. According to Variabilism, the loss Ann incurs at  $w_2$  bears on the permissibility of both  $a_1$  and  $a_2$  since Ann exists at  $w_2$ . In contrast, the loss Bob incurs at



w1 has no moral significance whatsoever, for purposes of evaluating either a1 or a2, since Bob never exists at w1. Given those moral data, the otherwise plausible permissibility theory will consider a1 obligatory and a2 wrong.

An advantage over Exclusion Alpha is that this account and these results will hold *whether or not it is w1 that is actual*. Suppose that w2 is actual and that a2 is actually performed instead of a1. The loss Bob incurs at w1 will remain, according to Variabilism, devoid of moral significance *despite the fact that Bob now actually exists*. We are thus never forced to say that the wrong act – that is, a2, the act that is wrong if left unperformed – becomes right if performed.

Variabilism has a similar advantage over Exclusion Beta. *Whether we happen to be evaluating a1 or a2*, the loss Bob incurs at w1 remains devoid of moral significance. Variabilism thus has no opportunity – provided, of course, that the permissibility theory is itself otherwise plausible – to generate *both* the result that a1 is obligatory *and* the result that a2 is permissible.

2.9.3 The conceptual advantage Variabilism has over Exclusion in both its forms is a function of the fact that, under Variabilism, the class of losses that have moral significance is *fixed*. The losses that bear on the permissibility of any one act or any alternative to that act do not shift depending on which act is actually performed or which act we happen to be evaluating. Variabilism thus achieves a certain stability, and consistency, across its various results.

Now, so does Inclusion. But Variabilism has clear advantages over Inclusion. Unlike Inclusion, Variabilism considers Bob's loss at w2 morally insignificant – and we accordingly avoid Inclusion's impression that a1 and a2 are nothing more than two peas in a moral pod. According to Variabilism, Ann *must* come before Bob; your actual dog *must* come before your merely possible cat. We thus find in Variabilism but not Inclusion the thesis that it *does* matter morally whether we *make people happy* or *make happy people*.

2.9.4 Variabilism also suggests a plausible account of Double Wrongful Life. According to Variabilism, the losses incurred by Cal at w1 and Dee at w2 *both* bear on the permissibility of *both* a1 and a2. If imposing on Dee the loss at w2 counts against a2, avoiding the loss Cal incurs at w1 counts, in a roundabout way, in favor of a2. If imposing on Cal the loss at w1 counts against a1, avoiding the loss Dee incurs at w2 counts, in a roundabout way, in favor of a1. Variabilism functions here just as Inclusion does – and parts ways with Exclusion. It appreciates a certain loss that Exclusion Alpha and Exclusion Beta blind themselves to – and can, accordingly, appreciate the parallels. The otherwise plausible permissibility theory will imply that a1 and a2 are both permissible.

2.9.5 Finally, we turn to Addition Plus. Let's first consider a1 – and compare it, first, against a2. According to Variabilism, the loss Etta incurs at w1 is morally significant. But so is the loss Fen incurs at w2 (compared to w3, of course, not w1). It's true that a2 avoids the morally significant loss Etta incurs at w1. But the morally significant loss Fen incurs at w2 is much deeper. Now – as noted before – just how this tradeoff between Fen and Etta is to be made is something we leave to the otherwise plausible permissibility theory. My own view is that a more complete version of that theory will say that the fact that a2 imposes the very great and very deep loss that it does

on Fen – compared, again, not to how well off Fen is at w1, but rather to how well off Fen is at w3 – means that a2 is in no position to establish that a1 is wrong.

In any case, what is very clear is that Variabilism, by understanding Fen's loss at w2 to *bear on* the permissibility of a1, parts ways with Exclusion Alpha (in the case where a1 is actually performed) and Exclusion Beta (since a1 is the act under scrutiny).

But Variabilism also parts ways with Inclusion. While Inclusion counts the loss that Fen incurs at w1 *against* a1 – making it that much harder to find a1 permissible – Variabilism instead treats that particular loss *as though it never happened*.

To complete the evaluation of a1, we must also compare a1 against a3. (So far, all we have done is to establish that a2 does not show that a1 is wrong.) According to Variabilism, the loss Etta incurs at w3 is morally significant since Etta exists at w3. But what about the loss that Fen incurs at w1? Here again, Variabilism parts ways with Inclusion – instead tracking Exclusion Alpha (in the case where a1 is actually performed) and Exclusion Beta (since it is a1 we are evaluating). It's true that a3 avoids the loss that Fen incurs at w1 and that a3 makes things better for Fen than a1 does. But there's a catch. According to Variabilism, since Fen *never exists at all* at w1, the loss Fen incurs at w1 has no moral significance. Thus, what we really have in a3 is an act that imposes a morally *significant* loss on Etta while avoiding a morally *insignificant* loss on behalf of Fen. To top it all off, a1 makes things better for Etta than a3 does. On those facts, there is nothing in the otherwise plausible permissibility theory that even begins to suggest that a1 is wrong.

But this is not yet to say that a1 is *permissible*. OPPP2 does not apply since a1 imposes a morally significant loss on Etta. OPPP3 does not apply since the changes are not merely reversing. Now, plausibly, if the principles within our relevantly complete permissibility theory do not imply that an act is wrong, we should think that that act is permissible. But we don't yet have a relevantly complete permissibility theory; this kind of inference is, accordingly, premature.

Indirectly, however, we can produce a firmer foundation for the result that a1 is permissible. And that is to rule out a2 and a3 – to find them wrong. Now, we won't be able to do this by reference just to OPPP1–OPPP4; we will have to speculate a bit how an otherwise plausible permissibility theory will direct that one particular tradeoff is to be decided. But once we do that, we are then done. For at that point standard deontic axioms will instruct that, if a2 and a3 are wrong, a1 is permissible.

Accordingly, we turn to a2 and a3. But this discussion will be brief, for, according to Variabilism, what counts as a morally significant loss does not vary depending on the act under scrutiny. So what we have said before about which losses are morally significant and which are not holds for the present discussion as well.

Let's take a2 first – and compare it against a3. Fen alone incurs a loss at w2 and Etta alone incurs a loss at w3. According to Variabilism, both of those losses are morally significant. We, accordingly, face a tradeoff. It is here that we can only speculate regarding the contents (beyond OPPP1–OPPP4) of the otherwise plausible permissibility theory. It seems plausible, however, to think that the tradeoff should be made in favor of Fen rather than Etta. In other words, the fact that a3 makes things better for Fen than a2 does, and thereby avoids a morally significant loss on his

behalf, all without imposing *too* deep a morally significant loss on Etta, plausibly means that a2 is wrong.

This analysis, reminiscent of Leximin, is somewhat controversial. Certainly, the Totalist and perhaps the Pluralist, while agreeing that Fen's loss at w2 is morally significant, will still find a2 obligatory. But a tradeoff principle that declares a2 wrong in light of the availability of a3 may seem more plausible.

We turn, finally, to a3. Here it's the comparison against a1 that is telling. a3 imposes a morally significant loss on Etta. a1 makes things better for Etta than a3 does. Now, it's true that Fen incurs losses at w1 (as compared to both w2 and w3). According to Variabilism, however, those losses have no moral significance since Fen never exists at w1. On these facts, the otherwise plausible permissibility theory – and, specifically, OPPP1 – will imply that a3 is wrong.

We conclude, then, that a2 and a3 are both wrong, and a1 is obligatory.

2.9.6 These are highly plausible, and consistent, results. We are seeing here the *variable moral significance* not just of *merely possible people* – Fen at w1 – but also of *existing or future people* – Fen at w2 and at w3. In some respects, the merely possible matter morally – just as, in some respects, do you and I. When we ask whether a2 shows that a1 is impermissible, what we see is that Fen, though merely possible relative to w1, matters morally just as much as Etta does. But in other respects, the merely possible do not matter at all – and nor do we. Thus, leaving Fen out of existence at w1 imposes, according to Variabilism, no morally significant loss on Fen. At that juncture of the analysis, Fen does not matter morally *at all*, not even the slightest bit, for purposes of evaluating a2 or a3 (or a1).

## 2.10 The Neutrality Intuition

2.10.1 Variabilism does not exhaust the middle ground between Exclusion and Inclusion. It shares that ground with some other views. One such view is suggested by what John Broome calls the *Neutrality Intuition*. Another is what Peter Singer calls the *Prior Existence View*. In certain respects, these two views are closely related. Each can be viewed as an attempt to capture our *person-affecting*, or *person-based*, intuitions. And each is in the end rejected – rightly, I think – by its respective articulator on the ground that it excludes from the analysis the interest of the person whose coming into existence is at stake.

2.10.2 According to the Neutrality Intuition, “adding a person [to the world] is equally as good as not adding her” – that is, “adding a person to the world is very often ethically neutral.”<sup>76</sup> Broome considers the Neutrality Intuition “part of the broader way of thinking known as the ‘person-affecting view.’ This is the view that

---

<sup>76</sup> Broome (2004), pp. 143 and 146.

only benefits or harms that come to people can be ethically significant.... A change is neutral unless it makes someone either better or worse off than she would otherwise have been.”<sup>77</sup>

Broome argues that we should reject the Neutrality Intuition. I agree. What I want to point out here is that his argument against the Neutrality Intuition does not apply to Variabilism. In fact, Variabilism provides a good account of just why it is that the Neutrality Intuition fails.

2.10.3 Broome’s first step in constructing an argument against the Neutrality Intuition is to state the view in a more precise way. He proposes the *Principle of Equal Existence*, which pertains, not to *act* permissibility, but rather to *outcome* betterness.<sup>78</sup> I paraphrase it here:

*Principle of Equal Existence:* If two outcomes, or *distributions*, include the same populations except that one contains an additional person whose life is in the “neutral range” – is, at least, not a life *less* than worth living – then those distributions are equally good *if and only if* they are equally good *for each person who does or will exist in both*.<sup>79</sup>

It is easy to see that the Principle of Equal Existence is an alternative to Exclusion and Inclusion. The Principle of Equal Existence does not require that people be actual in order to have their interests taken into account for purposes of ranking the distributions. The merely possible matter morally, too – as long as they exist in both the distributions to be ranked. At the same time, the principle excludes from the calculation the interests of the additional person even if that additional person is actual.

By implication, then, *leaving a person out of an existence worth having* is not going to make one distribution worse than the other, nor is *bringing a new person into an existence worth having* going to make the other distribution better than the one.

---

<sup>77</sup> Broome (2004), p. 145. Interestingly, the motivation that Broome cites for the person-based approach is just that bringing a person into existence “makes her neither better nor worse off.” That metaphysical point, he suggests, grounds the moral intuition for those who happen to have it. But we can also reject the metaphysical point – and I do; I think we can, and need to be able to, say that the person who never exists incurs a loss in the case where agents could have brought that same person into an existence worth having – and accept the person-based intuition on purely moral grounds.

<sup>78</sup> By *distribution*, Broome means to indicate a particular way in which wellbeing is distributed across a particular population. For purposes here, I have instead used the term *outcome*. (A world is different: two worlds may differ in many details but still generate the same outcome, or distribution – that is, the identical distribution of wellbeing across the identical population. And we still need worlds: we can talk about acts performed at worlds producing certain outcomes; but acts themselves can produce vastly different outcomes, depending on the worlds at which those acts are performed.) So for purposes here either term will do. The more material distinction is between outcome *betterness* and act *permissibility*. This distinction I address in the text. See Broome (2004), pp. 140–149; see also note 7 above (distinction between *deontic* and *teleological* project) and Appendix C.

<sup>79</sup> Broome (2004), p. 146. The Principle of Equal Existence is formulated in terms of the “neutral range,” and I assume here that the additional person whose life is “clearly worth living” has an overall wellbeing level that falls within that range. See also Broome (1991), p. 229 (discussion of the *Constituency Principle*).

2.10.4 As Broome shows, however, the Principle of Equal Existence leads to inconsistency when put together with still other principles we want to accept. Now, under the Principle of Equal Existence, it is *distributions* that are to be *ranked* in respect of their relative betterness, and not *acts* that are to be evaluated for their *permissibility*. With that caveat, we can display the sort of case that worries Broome in the usual way, with the alternative acts listed on the first row, and the distribution that is correlated in the case with each such act described below. As usual, the name in *italics* coupled with the asterisk means that the specified person never exists at all under the specified act, while boldface means that the specified person does, or will, exist under the specified act:

	a1 at w1	a2 at w2	a3 at w3	
+2	<b>Sam</b>	<b>Sam</b>	<b>Sam, Tim</b>	} Distributions
+1		<i>Tim</i>		
0	<i>Tim</i> *			

Case 2.4 Simple Addition

The Principle of Equal Existence implies that the distributions generated by a1 and a2 are equally good and so are the distributions generated by a1 and a3. The transitivity and symmetry of *equally as good as* implies that the distributions generated by a2 and a3 are also equally good. But we understand, on other grounds, that the distribution generated by a3 is clearly better than the distribution generated by a2.<sup>80</sup> So we have an inconsistency. We need to reject the Principle of Equal Existence.

2.10.5 Broome elsewhere suggests that coherent theories of act permissibility are easier to come by than are coherent theories of distribution goodness.<sup>81</sup> Interestingly, however, the root defect in the Principle of Equal Existence isn't at all mitigated if we extend that principle – and the idea behind it, that is, the Neutrality Intuition, the intuition Broome associates with the person-based approach – in a way that evaluates acts for their permissibility. Such an attempt will seem perfectly natural if we happen to accept – on the theory that the *consequences* of acts are critical for determining the *permissibility* of those acts – a strong connection between distribution goodness and act permissibility. What we will find, however, is that the results we obtain for act permissibility are just as problematic as they are for distribution goodness.

The case we've just examined – Simple Addition – makes this point. Suppose that a1 is itself permissible. Then, since the only difference between what a1 does and what a2 does is that a2 brings Tim into existence and accords to Tim an existence worth having, we should be able to infer that a2 is permissible as well. After

<sup>80</sup> Here Broome appeals to the *Principle of Personal Good*. Broome (2004), pp. 246–47.

<sup>81</sup> See Broome (2009), p. 248.

all, on the assumption that there is a strong connection between act permissibility and distribution goodness, if the distributions generated by *a1* and *a2* really are equally good, and if *a1* is permissible, then *a2* is permissible as well. For, if it were true that *a1* is permissible and *a2* wrong, then we should be able to see that the distribution generated by *a2* is morally worse than the distribution generated by *a1*. But we don't, according to the Principle of Equal Existence. By the same analysis, *a3* is permissible. But the fact that the distribution generated by *a3* is *better than* the distribution generated by *a2* should tell us that *a2* is impermissible. And we again we face inconsistency: *a2* can't be both *permissible* and *wrong*.

2.10.6 Broome's conclusion is that we must reject the Principle of Equal Existence. It doesn't work as a criterion for when one distribution is as good as another. Nor – as we have just seen – do things get any better when we extend the principle in a way that makes it relevant to the issue of act permissibility.

Variabilism provides us with a clear account of just why this principle and the Neutrality Intuition itself both fail. Let's stay, for the moment, on the topic of act permissibility. For that purpose, we need to retain the assumption that act permissibility and distribution goodness are strongly connected. Variability diverges from the extended Principle of Equal Existence when the latter claims that *a1* and *a2* are *equally good* and hence that, if *a1* is permissible, so is *a2*.

For it's at that point that the latter principle makes the mistake of focusing exclusively on what is good *for Sam*, who exists under both *a1* and *a2* and fares just as well under one act as he does under the other. It's at that point any *correct* principle should take into account the fact that *Tim incurs a loss under a2* – not, of course, a loss compared to how Tim fares at *a1*, but rather a loss compared to how Tim fares at *a3*. It's at that point that any correct principle will discern that loss and its relevance to the moral analysis. The approach to act permissibility that an extended Principle of Equal Existence suggests, however, blinds us to that very loss – or at least to its moral significance.

In contrast, Variabilism implies that the loss that Tim incurs at *a2* has *full moral significance*, for purposes of evaluating *a2* and each of its alternatives. After all, it is a loss Tim incurs at a world, *w2*, at which Tim exists. The otherwise plausible permissibility theory – the most standard of Pareto principles, and as well as the more restricted OPPP1, in combination with Variabilism – will quickly instruct that, given *a3*, *a2* is wrong.

2.10.7 Variabilism thus avoids the inconsistency that we derived when we tried to extend the Principle of Equal Existence in a way that makes it pertinent to act permissibility. Does an *extended Variabilism* – Variabilism extended, that is, in such a way that it addresses certain issues of distribution goodness – similarly avoid the inconsistency we derived from the original Principle of Equal Existence?

Again, retaining the assumption that act permissibility and outcome goodness are strongly connected, we can speculate as follows. If *a2* is wrong and *a1* is permissible – as Variabilism, in combination with the otherwise plausible permissibility theory, will instruct – then we want to say that it cannot also be the case that the distributions generated by *a1* and *a2* are equally good. We want to say that, from a moral point of view, the distribution generated by *a2* must be *worse than* the distribution generated by

a1 is. By a parallel analysis, we also can say that, since a2 is wrong and a3 is permissible, the distribution generated by a2 must be worse than the distribution generated by a3, and that, since a1 and a3 are both permissible, the distribution generated by a3 is exactly as good as the distribution generated by a1.

Intuitively, these results seem plausible. Moreover, there is no inconsistency.

We obtain the same results even if we don't rely on the results we obtain regarding permissibility for purposes of ranking the distributions in respect of their goodness. According to Variabilism, no one incurs any morally significant loss whatsoever under a1. In contrast, Tim incurs a fully morally significant loss under a2. Any otherwise plausible *ranking* theory may well instruct, on those facts, that the distribution that a2 generates – its consequence, that is – is morally worse than the distribution that a1 generates. Adding a new person – Tim – can thus be plausibly be considered to have *made things worse* from a moral point of view.

According to the Neutrality Intuition, “adding a person to the world is very often ethically neutral.”<sup>82</sup> But what the cases seem to suggest, and a view that Variabilism itself clearly supports, is that adding a new person to the world is very often *ethically hazardous*. This is not at all to say we shouldn't do it – in, for example, Simple Addition. It's just to say that if we do add a new person to the world, then we should do it by way of a3 and *not* a2. Who could disagree?

2.10.8 The Neutrality Intuition was proposed by Broome as a part of what he calls the person-based view.<sup>83</sup> Variabilism suggests an alternative understanding of the person-based intuition. We can use Simple Addition to illustrate the difference. The otherwise plausible permissibility theory itself – specifically, OPPP2 – implies that a2 is wrong *only if* a2 imposes a morally significant loss on some person or another. But this necessary condition on wrongdoing is consistent with an extremely wide person-based approach. It's consistent, for example, with Totalism. We don't get the *intuition* in the person-based intuition until we add a loss rule of a certain kind. Variabilism will certainly do for that purpose. Variabilism, in combination with OPPP2, thus implies that a2 is wrong at w2 *only if* a2 imposes a loss at w2 on a person who does or will exist at w2. And it seems that we should also be able to say that the distribution generated by a2 is *worse than* the distribution generated by a1 *only if* a2 imposes a loss at w2 on a person who does or will exist at w2. Thus:

a2 performed at w2 is wrong, and the distribution generated by a2 performed at w2 is worse than the distribution generated by a1 performed at w1, *only if* a2 imposes a loss at w2 on a person who does or will exist at w2;

What we have here is a simple necessary condition on wrongdoing: a way of understanding the intuition that the “bad” act must be “bad for” someone or another.<sup>84</sup> But it's a necessary condition that can be satisfied by Tim – for example – in the

<sup>82</sup> Broome (2004), p. 143.

<sup>83</sup> Broome (2004), p. 145.

<sup>84</sup> Parfit (1987), p. 363.



context of Simple Addition – since Tim does incur a loss at  $w_2$ . Thus: while Tim incurs no loss under  $a_2$  relative to  $a_1$ , Tim clearly does incur a loss under  $a_2$  relative to  $a_3$ . And that loss counts, for purposes of satisfying the necessary condition. The implication to the “no wrong done” result is, in other words, blocked.

In contrast, the necessary condition that the Neutrality Intuition establishes is more stringent. “A change is neutral unless it makes someone either better or worse off than she would otherwise have been.”<sup>85</sup> Thus, where *unless* means *if not*, and where we take the contrapositive, the Neutrality Intuition states that the change from  $a_1$  at  $w_1$  to  $a_2$  at  $w_2$  can make things worse *only if* someone – Tim, for example – is worse off at  $w_2$  than that person – Tim – is at  $w_1$ .

The distribution generated by  $a_2$  performed at  $w_2$  is worse than the distribution generated by  $a_1$  performed at  $w_1$  *only if*  $a_2$  makes things worse for a person who does or will exist at  $w_2$  than  $a_1$  makes things for that same person who does or will exist at  $w_1$ .

The Principle of Equal Existence makes that same point. But in the context of Simple Addition, this condition is not satisfied. Tim is, if anything, *better off* at  $w_2$  than he is at  $w_1$ . And Tim in any case exists only at  $w_2$ . And so we infer that the change from  $a_1$  to  $a_2$  can’t make things worse. That is:  $a_2$  is not worse than  $a_1$ . And that, of course, is just the inference that created the problem we noted above to begin with. But it’s a problem that Variabilism, in combination with the otherwise plausible permissibility theory, nicely avoids, since the focus for Variabilism is *not* whether the existing or future Tim is worse off at  $w_2$  *relative to*  $w_1$ , but rather whether Tim is worse off at  $w_2$  *relative to any other scenario whatsoever*.

One other note. In still another respect, the Neutrality Intuition is much broader than what we have in Variabilism in combination with the otherwise plausible permissibility theory. For the Neutrality Intuition also implies this:

The distribution generated by  $a_2$  performed at  $w_2$  is better than the distribution generated by  $a_1$  performed at  $w_1$  *only if*  $a_2$  makes things better for a person who does or will exist at  $w_2$  than  $a_1$  makes things for that same person who does or will exist at  $w_1$ .

The Principle of Equal Existence makes that same point. But this can’t be right. Given that it isn’t better for *anyone* that  $a_1$  be performed rather than  $a_2$ , these principles imply that  $a_1$  is no better than  $a_2$  is. And given that we know that  $a_2$  is wrong ( $a_3$  shows us that that is so), and a strong connection between distribution betterness and act permissibility, we must now say that  $a_1$  is wrong as well. Now, this result may not seem immediately objectionable in connection with Simple Addition. But precisely the same line of reasoning applies in the case of Addition Plus as well – forcing us now to say that  $a_1$  in that case is wrong: that is wrong *not* to bring additional merely possible persons into existence (e.g. Fen) and *treat them badly*. But we understand that that result cannot be right.

What we are seeing here is that it is a mistake to think that the simple *necessary* condition that the person-based intuition is meant to express can be swollen to try to express the necessary *and sufficient* condition that Broome associates with the

---

<sup>85</sup> Broome (2004), p. 145.

“person-affecting view.” The simple *necessary* condition on wrongdoing (understood by reference to Variabilism, not the Neutrality Intuition) is fine. But to convert that necessary condition on wrongdoing – on when one act is worse than another – into a *sufficient* condition is a *mistake*. In fact, it’s a logical error – something close to the fallacy of affirming the consequent – to turn this simple necessary condition around, add an essentially innocuous condition and then consider it something we are *compelled* to accept should we have happened already to have committed ourselves to the simple necessary condition.

Logic itself, then, does not require that move, and it is one that we should, moreover, avoid. It is easy to why that is so, if we now shift the case, from Simple Addition back to Addition Plus. We should reject – and logically are perfectly free to reject, even if we accept the *necessary* condition on wrongdoing – the following *sufficient* condition on wrongdoing:

a1 performed at w1 is wrong, and the distribution generated by a1 performed at w1 is worse than the distribution generated by a2 performed at w2, *if* a1 imposes a loss at w1 on a person who does or will exist at w1 and each person who does or will exist under a2 at w2 is at least as well off as that same person is under a1 at w1.

Addition Plus is an excellent counterexample against this variety of sufficient condition on wrongdoing. We think a1 is perfectly permissible, despite its effect on Etta. And – as we have already seen – it is a counterexample that Variabilism, in combination with the otherwise plausible permissibility theory, avoids.<sup>86</sup>

---

<sup>86</sup> For purposes here, I have assumed a strong connection between outcome betterness and act permissibility, and have freely gone back and forth from the one topic to the other. This approach clearly implies that the distributions generated by two acts can be ranked in respect of their betterness in one way if a third act producing still a third distribution is available to the agents, and in another way if it is not. The Axiom of the Independence of Irrelevant Alternatives could be taken to suggest that that cannot happen.

There are two points to be made in this connection. One is that there is no reason to think that the Independence Axiom does not apply nicely to the four-place betterness-for relation (the distribution generated by a1 at w1 is better for *p* than the distribution generated by a2 at w2 is for *q*) if not the two-place betterness relation (a1 is better than a2 *simpliciter*). The second is that, if we do think the Independence Axiom applies to the two-place relation as well, then it is plausible to think that what decides whether some a3 is an alternative for agents or not is in fact something inherent in a1 and a2. Thus: we might think natural law means that the world at which I am instantly cured of some terrible disease is not a world that is accessible to agents. But if that is so, then that is so because the world where I am sick keeps its agents firmly under the governance of that particular natural law. Taking that approach, when we specify (for purposes of determining permissibility) that a3 is an alternative to a1 and a2 for agents, and when we specify that a3 is not an alternative to a1 and a2 for agents, what we refer to by *a1* and *a2* itself shifts. In one case, there is something in the pair of options that precludes a3, and in the other case there isn’t. We can thus retain the Independence Axiom for the two-place betterness relation as well as the four-place betterness-for relation. I have discussed these issues elsewhere. See Roberts (1998) and Roberts (2003b). See also Appendix C below.

## 2.11 The Prior Existence View

2.11.1 Though a theory of act permissibility and not distribution betterness, what Singer calls the *Prior Existence View* is similar to the Neutrality Intuition in certain respects. The Prior Existence View states that, when evaluating whether to bring a new person into existence, we are to “count only beings who already exist, prior to the decision we are taking, or at least will exist independently of that decision.”<sup>87</sup> On this principle, we omit from our calculations the wellbeing the new person will accrue if the choice to bring that person into existence is made. That move, in turn, allows us to reject the highly counterintuitive idea that “the couple [should] count the likely future pleasure of their children as significant reason for having children....”<sup>88</sup>

2.11.2 As noted, Singer rejects the Prior Existence View. To see why, we need just consider the child “who, perhaps because it will inherit a genetic defect, would lead a thoroughly miserable life and die before its second birthday....”<sup>89</sup> Now, on some theories it will make a difference to the analysis just how miserable this “thoroughly miserable life” in fact is. The discussion can become very complicated in the case where the child’s life is, though miserable, unambiguously worth having.<sup>90</sup> But those complications are not important here. For purposes here, we can simplify the discussion by stipulating that the child’s life is *less* than worth living – that is, genuinely *wrongful* – if he or she exists at all.

According to the Prior Existence View, the fact that the child’s life will unavoidably be less than worth living must be excluded from the moral evaluation of the choice whether to bring that child into existence or not. But surely we can’t plausibly include that fact. “We would think it wrong for a couple

---

One last point on the Independence Axiom. If the approach I describe in this note turns out to be unworkable, then one option we have is to understand that the two-place betterness relation has no independent role to play in our moral analysis. We are very close to that result in any case. That is: the otherwise plausible permissibility theory relies, not on the two-place relation, but rather the four-place relation, to make its permissibility assessments. We can easily just understand the two-place relation as a short-hand way of saying that a particular act is wrong. To say, in other words, than an act *a* is better than an act *a'* – or, more precisely, that the distributions accordingly generated by *a* and *a'* rank in that particular way – is just another way of saying that *a'* is wrong.

See also Appendix C for discussion of an objection from Broome that targets, not the Neutrality Intuition as a way of articulating the person-based approach, but rather the person-based approach in combination with the idea that act permissibility and outcome betterness are themselves closely linked.

<sup>87</sup> Singer (1999), p. 103.

<sup>88</sup> Singer (1999), p. 103. Singer himself may anticipate that that calculation would be made on an aggregative basis if we accept the Prior Existence View. After all, Singer considers the Prior Existence View as an alternative to Totalism, whose aggregative feature Singer offers no objection to in this context.

<sup>89</sup> Singer (1999), p. 104.

<sup>90</sup> At that point, issues involving the nonidentity problem are triggered. We thus put that particular version of the case to the side for purposes here. But see part 2.2.5 above and Appendix B.

knowingly to conceive such a child. . . .”<sup>91</sup> At the very least, the fact that a person’s life will unavoidably be less than worth living *bears on* the permissibility of the choice to bring that person into existence to begin with. The Prior Existence View asserts that the one has nothing to do with the other. We should, accordingly, reject the Prior Existence View.

2.11.3 It may initially seem plausible that the one child’s “likely future pleasure” does not count *in favor of* the choice to bring that child into existence. But it is also very clear that the distinct child’s expected misery *counts against* the choice to bring that child into existence. But if the principle that justifies the first claim for us is that the person whose coming into existence is at stake does not *count* – if the principle that justifies the first claim for us is just the Prior Existence View – then we must not just accept the first claim but also reject the second claim. And that latter we cannot do.

Accordingly, Singer rejects the Prior Existence View. That much seems exactly right. On that basis, he then seems to accept a basic *symmetry* between happiness and misery.<sup>92</sup> That much is at least understandable. As we have just seen, we do consider the *misery* that a new person will predictably suffer to bear on the permissibility of the choice whether to bring that person into existence. So it seems that we must consider the *happiness* that the happy person can be expected to accrue to bear on permissibility as well. Misery counts against the one act under scrutiny, and happiness in favor of the other.<sup>93</sup>

Now, as a general, all-other-things-are-equal sort of proposition, that seems right. But we can discern in the pair of cases that we have at hand here a distinction that has nothing to do with that general proposition. When we say that the choice to leave the happy person out of existence is permissible, the happiness that is foregone under that choice is foregone by a person *who never exists at all*. And when we say that the choice to bring the miserable person into existence is clearly wrong, the misery that is incurred under that choice is misery that is incurred by a person *who does or will exist*.

With this distinction in mind, we can then use Variabilism to explain why happiness does not count in favor of the one act but misery does count against the other act. (We can explain, that is, what is called the *Asymmetry*.) It’s not that the interests of the new person are to be excluded across the board from our calculation of the permissibility of the choice to bring that person into existence. Rather, that person’s interests should be included, but on a *variable* basis. According to Variabilism, the loss the person incurs when that person is left out of an existence worth having has *no moral significance at all*. In contrast, according to Variabilism, the loss the person incurs when that person is brought into the miserable existence has *full moral significance*. The relevant distinction is not that one case involves happiness and the other misery, but rather that in one case the loss is incurred at a world where the

---

<sup>91</sup> Singer (1999), p. 104.

<sup>92</sup> Singer (1999), p. 104.

<sup>93</sup> See also McMahan (2009), pp. 54–60.

person never exists at all, and in the other the loss is incurred at a world where the person does or will exist.

We can thus (A) reject, with Singer, the Prior Existence View. We can also (B) accept as a general, all-other-things-are-equal sort of proposition – that is, subject to (C) following – that, just as creating additional happiness for a person bears on the issue of permissibility, so does creating additional misery for a person. But we can also (C) limit that general proposition in the way that Variabilism suggests. We can, that is, insist that (B) itself is subject to the principle that the loss that is implied whenever agents create less wellbeing for a person when they could have created more has *moral significance* for purposes of evaluating the choice that imposes that loss or any alternatives to that choice *if and only if* the person who incurs that loss does or will exist at the world at which that loss is incurred.

2.11.4 According to Variabilism, then, the Prior Existence View fails because it excludes from the permissibility analysis too many of the losses incurred by the person whose existence depends on how the choice under scrutiny is made. The same point holds for the Principle of Equal Existence – and, more generally, the Neutrality Intuition. Those principles fail as well because they exclude from the betterness analysis too many of the losses incurred by the person who exists under one but not the other of the two choices being ranked in accordance with their betterness.

In contrast, Variabilism insists that those losses do bear on the issue of permissibility – and betterness. It recognizes those very losses as morally significant for purposes of evaluating the choices that impose those losses and the alternatives to those choices that avoid those losses. Yet Variabilism recognizes those losses as morally significant without forcing us to jettison our intuition that the loss incurred by the person at a world where that person *never exists at all* is itself devoid of moral significance. We can still say that avoiding *that* loss, on behalf of that person, does not give us a reason to bring that person into existence. We can still insist, appealing to Variabilism, that such a loss *cannot* make the otherwise permissible choice that imposes that loss wrong.

## 2.12 The Asymmetry

2.12.1 Singer's discussion of the Prior Existence View takes him into the very deep waters of what is called the *Asymmetry*. Singer uses one-half of the Asymmetry to motivate the Prior Existence View – though feels forced by the other half of the Asymmetry in the end to reject the Prior Existence View.<sup>94</sup> Moreover, the Asymmetry is an important case in its own right – and an excellent test case for Variabilism.<sup>95</sup> So it is worth focusing on a bit more closely here.

<sup>94</sup> Singer (1999), pp. 103–105.

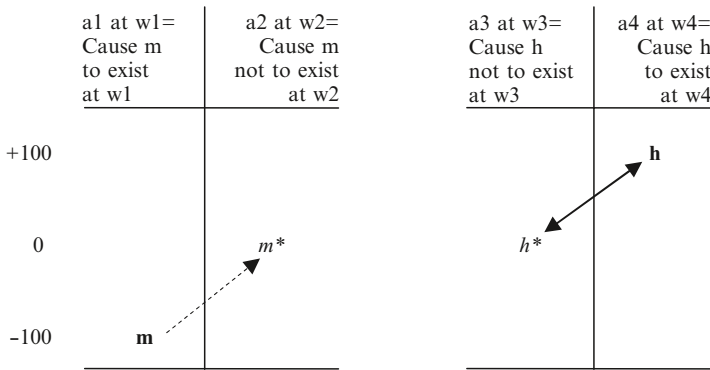
<sup>95</sup> See also Jeff McMahan, who gave a trenchant early description of the Asymmetry. McMahan (1981), p. 100, and Singer (1976), p. 92.

The Asymmetry comprises the following two claims:

*The Asymmetry.* Other things being equal, it would be wrong to bring a child into existence who would have a life less than one worth living, but it is permissible *not* to bring a child into existence even if that child would have a happy or even a wonderful life.

Singer and McMahan, as well as Ingmar Persson, recognize the commonsense appeal of the Asymmetry. But they all find it difficult to maintain. And each concludes, in the end, that we have no choice but to consider the additional person’s prospective happiness – in my terms, the *loss* that person will incur if that person is left out of existence altogether – as morally significant for purposes of evaluating the choice whether to bring that person into existence or not.<sup>96</sup> They each in the end conclude that the fact that the loss incurred by the merely possible person when that person is left out of existence altogether can make the otherwise permissible act that imposes that loss wrong. We are left to think that, after all, as Singer puts it, the “likely future pleasure” of one’s children is a “significant reason” for having them.<sup>97</sup>

2.12.2 We begin by graphing the pair of cases that make up the Asymmetry. Where *m* is the miserable child and *h* is the happy child, and the italicized name together with the asterisk signifies that the specified person never exists at all at the specified world and the boldface name signifies that the specified person does or will exist, the Asymmetry can be diagrammed as follows.



Case 2.5 The Asymmetry

It is the result that *a1* is wrong but *a3* is perfectly permissible – or, equivalently, that *a2* is obligatory but *a4* is not – that has seemed hard to explain. If *a2* is something that we *must* do, then why isn’t *a4* also something we *must* do? If *a1* is something we *can’t* do, why is *a3* perfectly permissible? If we have a strong moral reason for

<sup>96</sup> Singer (1999), p. 103. See also Persson (2009) and McMahan (2009).

<sup>97</sup> Singer (1999), p. 104.

leaving m out of existence altogether, why don't we also have a strong moral reason for bringing h into existence?

The problem is that many distinctions between a1 and a3, and between a2 and a4, seem incidental to the permissibility of those acts. As noted earlier, as a *general* proposition, we do think that just as misery can bear on permissibility, so can happiness. The idea that miserable children matter morally – and that accordingly any losses they incur have moral significance – but that happy children do not – and that accordingly their losses are somehow devoid of moral significance – is a non-starter. Moreover, the idea that losses below the zero level have moral significance whereas losses at or above the zero level do not seems arbitrary.

2.12.3 Exclusion Alpha and Exclusion Beta, of course, propose their own accounts in support of the Asymmetry. And the distinctions they rely on between a2 and a4 do intuitively seem to have *something* to do with why a2 is obligatory and a4 is not.

But those views come to the Asymmetry with their usual deficiencies fully intact. First, they only support our intuitions about half-way. Suppose a2 is actually performed. According to Exclusion Alpha, the loss m incurs at w1 has no moral significance at all. The otherwise plausible permissibility theory – specifically, OPPP2 – will then imply that a1 is permissible. Just as troubling, if a4 is actually performed, then Exclusion Alpha will say that the loss h incurs at w3 has full moral significance – and that a3 is wrong and a4, accordingly, obligatory. But both of these results are exactly contrary to what the Asymmetry says.

Exclusion Beta generates similarly problematic results whether it is a2 or a4 we happen to be evaluating.

Moreover, conceptual difficulties continue to plague both Exclusion Alpha and Exclusion Beta. Exclusion Alpha, in combination with the otherwise plausible permissibility theory, instructs that a1 is permissible if left unperformed – if, that is, a2 is actually performed – but instantly wrong if actually performed. a3 is permissible if actually performed – and instantly wrong if left unperformed. Exclusion Beta, still worse, instructs *for the same case* that a1 is wrong *and* a1 is permissible – and *for the same case* that a3 is permissible *and* a4 obligatory.

Inclusion simply denies the Asymmetry from the start.

2.12.4 Variabilism suggests a more plausible account of the Asymmetry. According to Variabilism, we do not decide whether a given loss has moral significance or not by tying that loss to a person and then trying to figure out whether that person matters morally or not – *counts* or not – for purposes of our evaluation. Instead, we recognize that *persons* themselves all matter morally in exactly the same way – in a way that is variable. It's their *losses* that do not all matter morally in exactly the same way. Some have moral significance, and some do not.

We need to consider, in particular, *where* the loss is incurred in relation to the person who incurs it. According to Variabilism, the loss incurred by either child m or h at a given world w bears on the permissibility of the act that imposes that loss and each alternative to that act that avoids that loss *if and only if* that child does or will exist at w.

Applying this loss rule, we divide the losses as follows. The loss m incurs at w1 has *full moral significance* for purposes of evaluating a1 and a2 since m exists at w1.



In contrast, the loss *h* incurs at *w*3 has *no moral significance whatsoever* – *not even the slightest bit* – for purposes of evaluating *a*3 or *a*4 since *h* never exists at all at *w*3.

It seems clear what the otherwise plausible permissibility theory will do with these moral data. On these data, any otherwise plausible permissibility theory – and, specifically, OPPP1 – will imply that *a*1 is wrong. The vast and morally significant loss *m* incurs at *w*1, in other words, where *m* does or will exist, makes *a*1 wrong, in light of the availability of the alternative *a*2, which is better for *m* and imposes no morally significant loss on anyone at all.

The same underlying analysis – specifically, OPPP2 – generates the result that *a*2 is permissible. Deontic axioms tell us still more – that, since *a*1 is wrong, *a*2 is in fact obligatory.

In dealing with the Asymmetry, the harder part has been to say what we have just said and *also say why a4 is not obligatory when a2 is* – that is, *why a3 is permissible*. According to Variabilism, however, what pushes us to say that *a*2 is obligatory – the *morally significant* loss that *m* incurs under *a*1 – is missing in the case of *a*4. For the loss that *h* incurs at *w*3, according to Variabilism, has *no moral significance whatsoever*. There thus is no morally significant loss that *a*4 avoids on behalf of *h*. The otherwise plausible permissibility theory, accordingly, implies that *a*3 is permissible – and so is *a*4. OPPP2 generates both those results. Variabilism thus agrees that *a*4 makes things better for *h* than *a*3 does; it agrees that *h* incurs a loss under *a*3 that *a*4 avoids on behalf of *h*. Variabilism simply asserts that that loss is devoid of moral significance (it is as though it never happened).<sup>98</sup>

Variabilism thus explains why the choice not to bring the happy child into existence is permissible while the choice to bring the miserable child into existence would be wrong. The difference is just that the loss for the miserable child takes place at a world where that real, live, flesh and blood child exists and suffers, whereas the loss for the happy child takes place at a world where that child never exists at all.

I suppose one might press harder on these two points. One might ask, for example, *why* the one loss matters morally and the other does not. But I think that we have said as much as there is to be said in response to that question. *It just does*. It's the difference between torturing your own real live dog and never bringing the Calico Cat or Thomasina or the Cat Who Went to Heaven into existence to begin with.

2.12.5 One last note in connection with the Asymmetry. The results that Totalism – as a paradigm example of the Inclusive approach – generates, regarding the first half of the Asymmetry, the half involving the miserable child, seem exactly right. But Totalism seems to go awry when applied to the second half of the Asymmetry, the half involving the happy child. It is there that we see that Totalism on the face of things seems to imply an implausibly stringent procreation obligation. We are

---

<sup>98</sup> These results parallel the results generated by the rights-based account that Persson explores. See Persson (2009). Persson himself rejects that account because of the difficulties inherent in the theory of rights that he is working with. I believe, however, that Variabilism, which is not rights-based, can track the plausible treatment of the Asymmetry that the rights-based account suggests while avoiding the difficulties inherent in that account.

alarmed! The Prior Existence View – which is a bit more exclusive – then comes to the fore. And the results that it generates for the second half of the Asymmetry seem exactly right. But then we apply it to the first half of the Asymmetry – and find that the results that it generates on that side are a disaster.

Variabilism arranges things so that we can take what we think is exactly right in a principled way from both Totalism and the Prior Existence View – and leave the rest. It fully supports the result that we do not have the stringent procreation obligation that Totalism itself suggests. That is, Variabilism, in combination with the otherwise plausible permissibility theory, flatly denies that we are required to perform a4. But it also flatly denies that that means that we can do whatever we want when it comes to bringing new persons into existence. Thus, in combination with any otherwise plausible permissibility theory, Variabilism insists that we are obligated not to bring the miserable child into existence – we are obligated, that is, *not* to perform a1. These are plausible results.

### 2.13 Summing Up

I thus believe Variabilism generates a plausible account of a significant range of problem cases. The loss incurred by the person who never exists at all cannot, according to Variabilism, make the otherwise permissible act that imposes that loss *wrong*. The accounts Variabilism provides of the Basic Case and the Asymmetry make that point. The implication for early abortion, and for non-conception and for conception and non-fruition, is clear: the option of an additional person coming into an existence worth having will not make the otherwise permissible choice to leave that person out of existence altogether wrong.

Just as important, Variabilism can say that without committing itself to the further result that what we have done is wrong in cases where *all we have done* is *decline* to bring a merely possible person into an avoidably and unjustifiably bad existence. That our choice *avoids* the maltreatment of the merely possible *can*, in other words, make the otherwise wrong act *perfectly permissible*. Double Wrongful Life, Addition Plus, Simple Addition and, again, the Asymmetry make that point.<sup>99</sup>

---

<sup>99</sup> I have proposed, in OPPP1 – OPPP4, some of the content of the otherwise plausible permissibility theory – which itself will consist of detailed principles that instruct, on the basis of the moral data that the loss rules under scrutiny will generate, which acts are permissible and which are not. We need, of course, to articulate – and test – more of that content. Appendix A (Otherwise Plausible Permissibility Theory + Variabilism) sums up what we have so far and suggests an extension.

Abortion and the Moral Significance of Merely Possible  
Persons

Finding Middle Ground in Hard Cases

Roberts, M.A.

2010, VIII, 190 p., Hardcover

ISBN: 978-90-481-3791-6