Modal Ethics

Part II: Narrow Neutrality

Melinda A. Roberts

robertsm@tcnj.edu

Chapter 7

Intuition and Existence

7.1 *Goals, organization of this Part II.*

7.2 *Terminology*. The terminology for this Part II is mainly the same as for Part I. We'll continue to define the term *person* broadly, connecting personhood with consciousness (including non-human consciousness) and understanding that the person who is *never conscious* at a given world *never exists* at that world.[1] We will continue to distinguish possible *worlds* (*futures*, *outcomes*) from *distributions*

---

[1] See Peter Singer. [*Animal Liberation*; *Practical Ethics*.] The term *person* thus includes many nonhuman animals and excludes many human beings. For purposes here, I assume consciousness to be a necessary but not a sufficient condition for a thing's being a person. And I assume that to survive as the same person from one time to another—for the person p at t1 to be numerically identical to the person q at t2—is for consciousness to be knitted together in some fashion or another by a transitive relation of psychological connectedness R. Moreover, I take it that a human or non-human embryo or fetus that hasn't experienced consciousness isn't a person; a human or non-human fetus that has experienced consciousness is in close proximity of, but isn't identical to, a person; and the person that may ultimately develop out of a human or non-human embryo or fetus doesn't come into existence until consciousness emerges. Thus: early abortion involves never bringing a person into existence to begin with whereas late abortion might (depending on facts about when consciousness emerges in humans) involve removing a person from existence.

and *possible* worlds from *accessible* worlds. The *Accessibility Axiom* will continue to play a role.

How *welfare* itself is precisely to be defined will remain, as before, an open question. But we can say that welfare indicates how good a person's existence at a given outcome is *for that person*. If a person p has more *welfare* in one outcome than p has in another, then the one outcome is better *for p* than the other. And we will continue sometimes to refer to the well-off person as simply the *happy* person.

Critical for this Part II will be the distinction between *welfare* (what is good *for the person*) and the *personal good* (how good the person's existence is *for the world*). If a person p has *more personal good* in one outcome than in another, then (other things equal) it will immediately follow in virtue of the meanings of the terms that that one outcome is *generally* better—is *overall* better, or *morally* better—than the other.[2] Room is thus left for the possibility that a person may have a positive *welfare* level in a given outcome even though that person's existence in that outcome contributes *nothing* to the *general* good of that outcome. That is, *welfare* may be positive even though *personal good* is zero.

---

[2] That is: on an additively separable basis. Thus if we say that the existence of a particular person at a personal good level of n at an outcome w1 contributes an amount n to the general good of w1, then we will say as well that the existence of that person at a personal good level of n at an outcome w2 shall also contribute an amount n to the general good of w2.

Chapter 8

Critique of Totalism

8.1 *Totalism*. Traditional consequentialist theories evaluate worlds on the basis of how much of that which makes life precious to the person who lives—how much *welfare*—those worlds contain. Such theories are *maximizing* in nature. A world that contains *more* welfare is *morally better* than a world that contains *less*.

Traditional consequentialist theories also take it as a given that no one person's welfare counts for any more than anyone else's does. Whether it's *your* child, or *my* child, who is assigned an extra unit of welfare is immaterial to whether one world, overall, is morally better than another. Such theories are *impartial*—non-agent relative, in other words—in nature.

Relatedly, traditional consequentialist theories are *inclusive* in nature. Your child, my child and everyone else who ever does, will or might exist has full moral status.

Traditional consequentialist theories, finally, make a *very tight connection* between the evaluation of *outcomes* and the evaluation of *acts*, where acts themselves are understood to include *omissions*. If an agent's act creates the *most welfare* that that agent can create—where the outcome an act produces is *morally better* than any alternative outcome—the act is *permissible*.[3] Otherwise, it's *wrong*.

Traditional consequentialist theories, so understood, seem to do a good job capturing the *basic maximizing intuition*—that idea, that is, that we ought to do the *best* we can—that is, create the *most welfare* that we can—for people. When we fall short of that—when we have created less welfare when we could have (other things equal) created more—what we have done is wrong.

The basic maximizing intuition makes sense to us. It seems right. However, the articulation of that intuition here is incomplete. For I haven't said what it is for one act to create *more welfare*, or for one outcome to contain *more welfare*, than

---

[3] This statement isn't quite right; an agent's *participation* in an act performed by a group may make what the agent has done wrong, even if the agent could not on his or her own have made things better. See Roberts ___ (on nip and collective action problem).

another. *More welfare* might, consistent with what we've said here, mean, for at least some person, more welfare *for that person*, or it might mean more welfare *in the aggregate*. Even so, the underlying intuition still seems right. But it won't follow that *any old way* of filling in that blank—that *any old theory* that *captures* that intuition but then goes on to say *a lot more as well*—is one we'll be compelled to accept.

Totalism, a paradigm example of a traditional consequentialist theory, fills in the blank by reference to *aggregate* welfare. Here, clause (i) sums up totalism's *telic* principle, and clause (ii) its *deontic* principle. Thus:

> *Totalism*:
>
> (i)   Where p is any person and W(p, w) is p's individual welfare level in a world w,
>
> *Total good* (w) = $\sum$ W(p, w) for each person p who ever exists in w;
>
> and
>
> wα is *morally better* than wβ iff total good (wα) > total good (wβ);
>
> and
>
> (ii)   An act aα performed at a world wα is *obligatory* at a time for an agent iff, for each wβ accessible at that time to that agent such that there exists no accessible wγ that is morally better than wβ, that agent performs aα at wβ.[4]

According, then, to totalism, the agent's *moral obligation* is just to perform that act that we find in the *morally best* of all those worlds available, or *accessible*, to the agent at a given time, where one world is better than another just in case the *total*,

---

[4] The principle of moral obligation set forth in the text is based on Fred Feldman, *Doing the Best We Can* (1986). For principles governing moral permissibility and conditional obligation, also see Feldman.

or *aggregate*, of all the welfare that world creates for each person who does or will exist in that world is greater.

In contrast, *person-affecting* views fill in the blank by reference to *individual* welfare. We'll defer until later the issue of just how such a view is to be articulated. But—as we shall see—it's very easy to come up with a view that is roughly *person-affecting* in nature but is clearly false.

8.2 *Objections*. I think the articulation of the basic maximizing intuition set forth above starts off well. The question is whether—in the hands of the totalist—it ends well. The connection totalism makes between an act's, or a world's, being *better* and the act's generating, or the world's containing, more welfare *in the aggregate* has to give us pause.

For one thing, it's the fact that totalism is aggregative in nature that means that totalism rules out the *happy child* half of the procreative asymmetry—the intuition, that is, that the existence of an additional *happy* child doesn't make a world better and that, other things equal, it's perfectly permissible not to bring that child into existence.[5] Creating more welfare *in the aggregate* is something that totalism obligates agents to do. Hence bringing the happy child into existence is something that totalism obligates agents to do. If the focus instead were, for each person, creating more welfare *for that person*, and then drawing a distinction between creating more welfare by way of bringing that (possible) person into

---

[5] This is not to say aggregation *on its own* rules out the happy child half of the asymmetry. Rather, it's to say that aggregation in combination with totalism's *other* features—including its maximizing feature and, relatedly, its unrestricted inclusion, for purposes of evaluating a given world, of the welfare level of *all* people, *each and every one* of them, who does or will exist at that world—that rules out the happy child half of the asymmetry.

Moreover, a non-aggregative (or "person-affecting") articulation of the basic maximizing intuition might—depending on its details—itself rule out the happy child half of the asymmetry. Consider, e.g., a theory that draws no distinction between the welfare created by way of bringing a happy child into existence and the welfare created by way of making an existing or future child better off.

existence and creating more welfare by way of making that (existing or future) person better off, the picture would be quite different.[6]

The basic maximizing intuition is very strong. But so is the happy child half of the asymmetry. Thus we have to ask: do we have two *intuitions* that are in conflict—the basic maximizing intuition and the happy child half of the asymmetry? If so, that's one strike against intuition. Or, alternatively, does totalism simply offer an imperfect—really a quite *bad*—reading of the basic maximizing intuition? If so, that's one strike against totalism.

Totalism's aggregative feature doesn't just put totalism at odds with the happy child half of the asymmetry. It also means that totalism is at odds with what we think are the right things to say about the repugnant conclusion, replaceability and the infinite population problem. [Brief description of each problem to come.] These Parfitian difficulties are, of course, in addition to a handful of perennial—but just as deep—objections to totalism, including objections based on equality, fairness and priority.

8.3 *Plusses of totalism.* It may seem that we already know quite enough to simply reject totalism—and, more generally, aggregation—outright.

But eliminating *all* forms of aggregation from our moral theory isn't really such a simple matter. Nor is it clearly desirable. Aggregation—or summation, or addition—has its plusses.

For one thing, aggregation may seem conceptually necessary to the basic maximizing idea itself. If *more* welfare is morally better than less, doesn't it just follow as a conceptual matter that more welfare *in the aggregate* is better than less?[7]

---

[6] This is not to say there is a moral distinction between existing and future people and merely possible people. All people, in my view, have the same moral status. But it doesn't follow that all ways of creating additional welfare for a given person have the same moral status. See Roberts [asymmetry papers].

[7] I have argued elsewhere that it doesn't. See Roberts 2002.

For another, totalism matter of factly, nicely, generates the *miserable* child half of the asymmetry. Totalism thus instructs that, other things equal, the existence of a *miserable* child—the child whose existence is *less* than worth having; the child whose life is *wrongful*—makes an outcome *worse* and bringing the *miserable* child into existence is *wrong*.

That totalism generates the miserable child half of the asymmetry is a reflection of the fact that totalism is aggregative in nature in combination with its being maximizing and inclusive in nature. Those three features together insure that totalism aggregates *without restriction* across the entire population at each and every world subject to comparison—and hence that totalism doesn't cordon off the misery of the miserable child as somehow lacking in moral significance.[8] The misery of the miserable child has full moral significance according to totalism, just as the happiness of the happy child has full moral significance according to totalism. And that's so, *even though* the miserable child's very existence is what is stake; *even though* the miserable child need not ever exist at all; and *even though* the miserable child exists in only one, not both, of the outcomes under scrutiny.

Appreciation of how totalism handles the miserable child half of the asymmetry softens what otherwise might seem a clear affront to intuition. We *thought* the happy child half of the asymmetry was one of our strongly held intuitions. But now we see—that is, we might *seem* to see—that letting go of that

---

[8] If the theory cordoned off the misery of the miserable child as somehow lacking in moral significance, then theory couldn't also be *fully maximizing* in nature: one world could turn out to be better than the other merely as a function of its having ignored the plight of the miserable child.

More generally, it might seem that we could preserve the happy child half of the asymmetry while retaining aggregation by simply restricting the scope of those individual persons whose welfare levels matter for purposes of aggregation. It might seem, e.g., that we could simply say that the happy child's welfare level is outside the scope of the aggregative function in virtue of the fact that that child exists in one but not the other of the two worlds that are the subject of our comparison. But it's widely recognized at this point—by Singer; Arrhenius; and others—that that restriction fails. As Singer argues, that sort of approach would compel us to reject the miserable half of the asymmetry, something we are loathe to do.

intuition is the perfectly reasonable price we must pay to retain the miserable child half of the asymmetry.

Relatedly, totalism generates *stable* results. Suppose that in the end the choice is made *not* to bring the miserable child into existence. That fact doesn't, according to totalism, somehow render the *nonactual* world better than it would have been had it been *actual*, or the unperformed choice to bring that child into existence *permissible*. The misery of the miserable child has moral significance, according to totalism, whether that child ever exists or not. Consequently, our moral evaluation of the act of bringing the miserable child into existence does not shift depending on whether that act happens in the end to have been performed or not.[9]

And there is more. The fact that totalism is aggregative in nature comes with many theoretical advantages. [[[It allows for easy theoretical check for transitivity; convenience of pairwise comparison.]]]

8.4 *Can we retain the happy child half of the asymmetry?* It might seem that any theory that displays the virtues we've just associated with totalism—and specifically with the fact that totalism is maximizing, aggregative and inclusive in nature—will immediately rule out the happy child half of the asymmetry.

John Broome, however, explores whether that's in fact so. Though he doesn't put things (or perhaps even, perhaps, conceive things) in this way, his discussion of what he calls the *neutrality intuition* can be viewed as a discussion of whether totalism can be corrected in a way that preserves the happy child half of the asymmetry *without* abandoning aggregation.

Thus Broome argues that the neutrality intuition leads to inconsistency. But—as we shall see—his argument against the intuition makes no reference to aggregation. Rather, the principles that Broome relies on to show inconsistency

---

[9] Hence no violation of Rabinowicz's Principle of Normative Invariance.

are, instead, principles that find wide acceptance by aggregationists and non-aggregationists alike.

My point then is just that *even the aggregationist* might have an interest in whether the neutrality intuition can be made to work—should such a theorist want to retain the happy child half of the asymmetry *without* jettisoning aggregation. The point is worth mentioning here since—as we shall also see—we might well want to locate ourselves in just that camp.

But for the moment to the topic is Broome. Many philosophers have found his inconsistency argument against the neutrality intuition compelling. I, too, find it compelling—at least given a certain restriction. I think it clearly shows that the neutrality intuition—in the form of what we shall call the *neutral range* claim, and subject to the same restriction—must go. But we can—and will—nonetheless question is whether the happy child half of the asymmetry must go as well.

Chapter 9

Correction to Totalism

9.1 *The neutral range claim.* We can think of the *neutrality intuition* as an attempt to correct a defect in totalism—or at least as an attempt to take seriously *other* philosophers' concerns that totalism is in need of correction.[10]

As Broome articulates it, the neutrality intuition states that there is a *neutral range* of existence such that a person's existing within that range in a given outcome does not, other things equal, make that outcome morally better or worse but is rather *neutral* in its effect. The neutrality intuition, as Broome articulates it, is really just what we can call the *neutral range claim.*

Thus he writes that "for a wide range of lives the child might live, having a child seems an ethically neutral matter."[11] And: "[T]here is some range of wellbeings (called 'the neutral range') such that, if the extra person's wellbeing is within this range, the two distributions are equally good," where the term range is meant to indicate "more than one member; the idea applied for several different levels of wellbeing."[12]

Broome restricts the neutrality intuition—that is, the *neutral range claim*—to cases in which the child's existence falls into the "neutral range." I shall assume that that range is meant to include the very cases of interest to us in the context of our consideration of the happy child half of the asymmetry—that is, cases in which the child's existence is *unambiguously* worth having; cases in which the child's existence *isn't* marginal and *isn't* just *barely* worth having.

On that assumption, the neutral range claim easily generates the result that, other things equal, the outcome in which the happy child exists *isn't morally better* than the outcome in which the child never exists.

---

[10] [Thus Broome cites Narveson. I, too, cling to the happy child half of the asymmetry, as does Heyd.]

[11] Broome 2004, p. 144.

[12] Broome 2004, p. 146.

Though Broome sometimes makes reference to the evaluation of *acts*—when he notes, for example, that there seem to be cases in which "having a child seems an ethically neutral matter"—he doesn't as a matter of theory accept the very tight connection between the evaluation of outcomes and acts. But if we do find that connection plausible—and I do; I am otherwise unclear what the purpose of determining whether one outcome is *morally better*, or even *generally*, or *overall*, better, than another is—we can see that the neutral range claim also generates the result that, other things equal, the agent's not bringing the happy child into existence is *perfectly permissible*.

Restricting the scope of the neutral range claim to existences that fall into the neutral range enables that claim to support the happy child half of the asymmetry *without* denying the miserable child half.[13] Thus we can say that the existence of the miserable child—the existence that is *less* than worth having—falls *outside* the neutral range. As such, the existence of the *miserable* child isn't caught by the neutrality intuition. And we thus are free, on other grounds, to say that the existence of the miserable child makes the outcome *worse* and that the agent's bringing that child into existence is *wrong*.

Just then to note: nothing we have said so far about the neutral range claim requires us to *abandon aggregation* in order to accept the intuition. What we must do, instead, is abandon a *fully inclusive* commitment to maximization.[14]

Broome notes that he himself finds the neutral range claim attractive. But he argues that in the end we must reject it as inconsistent.

9.2 *Narrow neutrality.* I find Broome's argument against the neutral range claim compelling. But I will argue in what follows that we can accept Broome's argument but at the same time recognize an intuition *behind* the intuition—an

---

[13] Thus it isn't *coming into existence* per se that we set aside (as Heyd sometimes seems to do) as having no effect on how good, overall, an outcome is. Rather it's the coming into existence *within a certain range* that has no effective.

[14] See note [8] above [fully maximizing in nature].

intuition I will call *narrow neutrality*. I will, in other words, argue that Broome's argument against the neutral range claim gives us *no reason at all* to think that we cannot accept narrow neutrality if that in the end is what we want to do.

But I have a further purpose as well. I will argue that we can retain narrow neutrality *without* setting aside all semblance of the aggregative function that, alongside maximization and inclusion, are responsible for the many plusses that come with totalism.

Specifically, I will argue that we can retain narrow neutrality consistent with the very principle that Broome himself deploys in order to rescue an additive approach from a number of traditional objections against totalism, including those based on the values of *equality*, *fairness* and, perhaps, *priority*. Thus Broome argues that his own formulation of Harsanyi's theorem—what I will call *P** in what follows—avoids just such objections. My argument will be that there is no reason to think that P* cannot be understood to avoid objections based on our *existential* values as well. If the *egalitarian*, or the *prioritarian*, can accept P*, so can the *narrow neutralist*, that is, the *existentialist*, accept P*.[15]

Now, it may seem that the additive P*—and, more generally, that any additive, or aggregative, approach—will force on us all the things that concern us about totalism—all the things, that is, that we *don't* like about traditional forms of consequentialism. P* is undeniably aggregative in nature: it aggregates by way of simple summation across the population of a given outcome to determine whether that outcome is overall (Broome says *generally*; I would say *morally*) better or worse than another. Thus, just like totalism, P* may seem not to "take seriously

---

[15] Broome presents P* as his own interpretation of Harsanyi's theorem. See GPG. Others refer to it as *additive separability*. The idea is that the good each additional person's existence contributes to the overall good of the outcome is independent of facts about the good other people's existences contribute to the overall good of that outcome. Specifically, the good the additional person's existence contributes to a given outcome is not deflated by the fact that the average good existence contributed by others who do or will exist at that outcome is higher, nor is it deflated by the fact that the number of well-off people who do or will exist at that outcome exceeds a certain level.

the distinctions between people."[16] And it may thus seem immediately to rule out considerations of equality, fairness and priority—right along with the happy child half of the asymmetry.

In fact, however, P* isn't so closely tied to totalism. Where totalism deploys the unadorned concept of *welfare*, P* instead puts the *highly* adorned concept of the *personal good* to work. It's that fact, Broome argues, that turns P* into a far more defensible principle—a principle capable of recognizing a myriad of values that totalism itself is completely oblivious to.

At the same time, in part as a function of the fact that it is additive in nature, P* has the very plusses we earlier attributed to totalism—the plusses, that is, but not the minuses. [It thus accommodates the miserable child half of the asymmetry; it's a straightforward way of articulating the basic maximizing intuition; it's inclusive; it's results are stable; it allows for pair-wise comparisons between outcomes and it helps us check our work—check for, e.g., failures of transitivity.]

One note. As we go about fitting narrow neutrality into a framework that includes P* but avoids Broome's inconsistency argument, we shall discover that an *inversion* of the calculation of the *personal good* from that which Broome himself may have had in mind in order. Inversion will be critical to understanding just how P* itself can account for narrow neutrality. But inversion will also help us explain the deeply held intuitions we have in connection with some of the other many problem cases in population ethics as well. Or so I will argue in what follows.

---

[16] Rawls, *A Theory of Justice*.

Chapter 10

The Neutral Range Claim

10.1 *Tradeoff.* We've already noted that totalism is at odds with the happy child half of the procreative asymmetry. The neutral range claim tries to correct totalism in a way that preserves the happy child half of the asymmetry without forcing us to reject the miserable child half of the asymmetry.

For both halves of the asymmetry, it's part of the case that the existence of the additional child affects no one else. But the defect in totalism that is at play in its treatment of the asymmetry comes to the surface even more clearly, I think, in cases in which the existence of the additional child *does* affect others. So let's start by noting how totalism fails in a case of that sort—I'll call it the *tradeoff case*—and how the neutral range claim seems initially to help.

The tradeoff case involves just two options: bringing a happy person into existence by way of imposing a steep decline in welfare for a distinct person and avoiding that steep decline in welfare on behalf of that distinct person by way of leaving the happy person out of existence altogether. It's immaterial to the case whether that distinct person is an *already-existing*, or a *future*, person.

The outcomes displayed in Graph __ (i) exist as *accessible* outcomes in the case and (ii) *exhaust* those outcomes. Bold face means the indicated person does or will exist in the indicated outcome, and italics paired with the "*" means the indicated person never exists in indicated outcome.

Graph __: Tradeoff	Welfare	w1	w2
Life well worth living	+10	**George**	**Jill**
	+9 … +2		
Life barely worth living	+1		**George**
	+0	*Jill**	

We are to suppose here that George's life at +10 goes really well for him in w1 and he is considerably worse off, at +1, in w2. If and only if George's welfare is reduced from +10 to +1, Jill will exist and have a life in w2 that at +11 is a little better than George's life is in w1. Total welfare being greater in w2 than in w1, totalism immediately implies that w2 is morally better than w1 is—and that it would be wrong to protect George at Jill's expense.

But both the telic and the deontic results here seem false. If we agree that that's so—and my aim here is not to argue that it is but rather to query whether we can consistently take the position that it is within a framework that otherwise seems plausible to us—then we will consider the tradeoff case to represent a serious problem for the totalist.

10.2 *How the neutral range claim helps.* Totalism implies that the two ways of adding welfare represented in the tradeoff case—adding welfare to Jill's stock and adding welfare to George's—work equally well.

The neutrality intuition—in the form of the neutral range claim—comes along and says that that's a mistake. It's a mistake to see Jill's welfare in w2 as adding to the total good of w2. Rather, Jill's existence in w2, despite her relatively

high welfare level, should be counted as morally *neutral*—as an addition that doesn't make the outcome better or worse.[17] We understand, on other grounds, that the effect on George of bringing Jill into existence isn't *neutral* at all—that what is done to George in w2 *does* make w2 worse. And we can then see how a perfectly routine account of the case would proceed to get us to the results that w2 is *overall* worse than, not better than, w1 and that bringing Jill into existence at George's expense is wrong.[18]

Those results seem entirely plausible. When the tradeoff is between bringing one person into existence and avoiding a loss on behalf of an existing or future person, it's the latter, not the former, that makes things better.

10.3 *Broome's inconsistency argument*. The neutral range claim seems to offer just the sort of intuitive correction totalism needs. Broome argues, however, that the neutral range claim is inconsistent. Consider the following *three outcome case*. As before, we stipulate that the displayed outcomes all three exist as *accessible* outcomes within the particular case. I should go ahead and note that that restriction—the *intra-case* restriction, I'll call it—is one that Broome himself disputes. Nonetheless, we'll first work through the argument with that restriction in place. For it's that form of the argument that—I believe—tells us something important about the neutrality intuition. It tells us the neutrality intuition, in the form of the *neutral range claim* and subject to that restriction, is inconsistent—and it suggests a better way of articulating the underlying intuition—the intuition *behind* the intuition, that is, *narrow neutrality*. We'll then consider how the argument unfolds *without* the restriction in place. There, I'll make the case that the argument fails.

---

[17] All we need to assume here is that Jill's welfare in w2 falls into the neutral range; that George's welfare in w2 might fall *below* that range is incidental to how the neutrality intuition applies since it isn't George's existence that is at stake.

[18] I won't delay things by laying out the specific principles here but I think they are both obvious and highly plausible. But see Roberts [Abortion and the MS of MPP].

Graph __: Three Outcome Case	Welfare	w1	w2	w3
	+10			**Paula**
	+5		**Paula**	
	+0	*Paula**		

It's an assumption of the case that Paula's existence in both w2 and w3 falls into the neutral range. Let's suppose that she has a very good life in w2 and an even better life in w3. We then compare w2 against w1. According to the neutral range claim, Paula's existence in w2 is neutral—it doesn't make w2 either better than or worse than w1. It follows, given the simplicity of the case, that w2 and w1 are equally good.[19] The neutral range claim produces parallel results when we turn to compare w3 against w1: w3 and w1 are equally good. Assuming that the *equally as good as* relation between outcomes is both transitive and symmetrical, we then infer that w2 is equally as good as w3. But we understand, on other grounds, that w2 is worse than w3 is. Given that w2 and w3 both exist as accessible outcomes per the intra-case restriction, the principle that moves the argument forward can be understood to be a simple, straightforward Pareto-like principle: where two such accessible outcomes contain exactly the same people, and one outcome is better for at least one person and worse for none than the other outcome is, then the one

---

[19]   I don't see the cases at issue in either Part I or Part II as challenging the completeness of the betterness relation. If X isn't better than Y and Y isn't better than X, then X is equally as good as Y is. I leave aside the question whether more complicated cases may represent legitimate challenges to completeness. [Cite R. Chang.]

outcome is itself worse. Hence we have an inconsistency. w3 can't be equally as good as *and* better than w2 is.[20]

Conceding both the transitivity and symmetry of the *equally as good as* relation and the claim that w2 is worse than w3 is, to avoid inconsistency we are forced to reject one or both of our two neutrality results. We must then reject either the result that w2 and w1 are equally good or the result that w3 and w1 are equally good. To reject either one or both those results is itself, of course, to reject the neutral range claim. So we reject the neutral range claim.

Notably, however, what we *can't* validly derive from Broome's argument is that *both* disjuncts are false. The inconsistency shows we must reject *one or the other* of the two disjuncts. But it doesn't show that we must reject *both*.

For example, we can avoid inconsistency by claiming that w2 and w1 *aren't* equally good. Specifically we can say that w2 is *worse* than w1. We are then free

---

[20] Broome 2014, pp. 146-147. Broome formulates the inconsistency argument not in terms of *welfare* but rather in terms of what he calls *wellbeing*. Now, his use of the term *wellbeing* in 2004 is at least at some points arguably synonymous with what he in 2015 calls the *personal good*. As we shall see, the *personal good* is itself an amazingly accommodating, *highly* adorned concept. The problem is that if his inconsistency argument is meant to make use of that *latter* concept—that is, the concept expressed by *wellbeing* in 2015— then we can't even set up the case for purposes of testing the neutral range claim without tripping over our own terms. Thus, *by definition*, the existence of a person in an outcome at a positive level of the *personal* good *increases* the *general* good. Broome 2015. But that means that if, in the 2004 inconsistency argument, we take +10 in w3 refers not to welfare but rather to the personal good, then Paula's existence at +10 in w3 would *immediately* generate the result that w3 is better than w1, a result that would in turn *automatically* rule out the neutral range claim. Broome would then have no need to draw on the simple, straightforward Pareto-like principle to show inconsistency. But he clearly *does* draw on *some* version of that very principle. (Now, just *which* version he means to draw on will be up for discussion later. See part ___ below. But for now the important point is that he draws on *some* such principle.) Hence it seems we should understand the inconsistency argument to refer not to the *personal good* but to the unadorned *welfare* instead.

I suspect that is indeed just how Broome means *wellbeing* to be construed in this particular context. And that seems so, despite the fact that things are further confused by Broome's own 2004 name for the Pareto-like principle: the principle of personal good. Broome 2004, p. 120.

to insist that w3 and w1 *are* equally good, that is, that Paula's existence in w3 as compared against w1 *is* morally neutral.

Broome's argument is thus *not*—at least not *immediately*, on its face, without further supplement—an argument against the claim that there is *some* level of welfare such that Paula can be brought into existence at that welfare level in that outcome without making that outcome either better or worse than w1 is. Rather, it's an argument against the claim that there exist *two or more* such levels.

At points, Broome himself restricts the conclusion he draws from the inconsistency argument in exactly that way. Thus he says that the argument tells us that there exists *at most* a *single neutral level*, a "sharp boundary," in the three outcome case—at most a *single* level of welfare such that bringing Paula into existence at that level does not make things morally better or worse.[21]

But that result conforms *perfectly* to what we might well consider the intuition *behind* the neutrality intuition—the intuition I will call *narrow neutrality*. We can thus *easily* let go of the idea—derived from the neutral *range* claim—that bringing Paula into an existence that makes things *worse* for her when things could have been *better* is morally *neutral*. We can quite happily instead say that Paula's existing at a *avoidably lesser existence* in w2 makes w2 *worse* than w1 is. We can quite happily instead say that, given w3, w2 is worse than w1 is, which is just to say that w3 shows that w2 is worse than w1 is.

But the moment we agree that w2 is worse than w1 we avoid the inconsistency while giving ourselves the option of retaining narrow neutrality. We thus can say that, in the particular case and for the particular person, Paula, it's in w3, not w2, where the "sharp boundary" of the neutral level is itself achieved. According to *narrow neutrality*, it's at that level and at that level alone that Paula's existence is morally neutral.

We can thus say about the three outcome case exactly what I think we *want* to say. Though Paula's existence in w2 makes w2 *worse* than w1 is, her existence in w3 *doesn't* make w3 *better* than w1 is.

---

[21] Broome 2004, p. 142.

A critical point. None of what I have said so far indicates, for purposes of understanding narrow neutrality, how the *neutral level* is itself to be defined. We can, however, note that it's not plausible to say that the neutral level is, e.g., always +10. If our facts were just slightly different—if Paula's welfare in w3 is not +10 but rather +9—we would still want to say that her existence in w2 makes w2 worse than w1 but her existence in w3 is neutral. Or if the case includes still a fourth accessible outcome—a w4 just like w3 except that Paula's welfare in w4 is +11 rather than +10—we would then want to say that her existence in w3 isn't after all neutral. We would want to say, that is, that w4 shows that Paula's existence in w3 makes w3 worse than w1. Thus what counts as the neutral level will be *case-*, or *context-, dependent*. Nor, for reasons having to do with cases in which what is at stake is the existence of two or more people and the changes between one outcome and the other constitute *merely reversing changes* (Vallentyne), do we want to say that the neutral level is the *maximal* level welfare that might be achieved for a given person within a given case. So there is no simple formula for calculating the neutral level for a given person in a given case. We shall thus need to come back to this question. But that the neutral level isn't rigidly fixed for all people and for all cases doesn't mean that it doesn't exist.

10.4 *Interpreting the argument*. Let's go back to the original three outcome case and Broome's inconsistency argument. Would Broome concede that that argument against the neutral range claim *doesn't* rule out the position that w3 and w1 are equally good? That it doesn't, that is, rule out *narrow* neutrality? It seems that he surely would have acknowledged that his argument opens the door to the position w3 and w1 are equally good if he thought that it did. Moreover, there is some reason to think that Broome might have meant for his inconsistency argument to rule out from the start the position that w3 and w1 are equally good. We consider both sides of the question here.

Consider how Broome introduces the neutrality intuition.

> *Neutrality intuition*: "Adding a person to the world is very often ethically neutral."[22] And, quoting Narveson, "'we are . . . neutral about making happy people.'"[23]

Immediately we have a question. "Often"? What's "often"?

The locution "often" might be meant just to recognize the *miserable child* exception to the neutrality intuition—to note, that is, that cases where the person's existence falls *below* the *neutral range* are outside the intuition.

But it's also possible that Broome's "often" is meant to recognize exceptions *beyond* the miserable child exception. It's possible, that is, that Broome's "often" is also meant to recognize an exception to neutrality Narveson himself would likely approve—that is, the *avoidably lesser existence* exception. If so, then Paula's existence at w2 is (like the miserable child's existence) would fall outside the intuition.

On this reading, the neutrality intuition *doesn't* imply that Paula's existence at w2 is neutral but rather implies just that existence is neutral often *enough—enough* being in the case at hand just *once*, that is, Paula's existence in w3—to instruct that w3 isn't morally better than w1 is.

That would mean, in turn, that the inconsistency argument itself targets, not the claim that w2 is equally as good as w1 *and* w3 is equally as good as w1, but rather just the claim that w3 is equally as good as w1 is.

That reading of Broome may seem at odds with what he says about the neutrality intuition. Thus he explicitly notes that he uses the term *range* to "imply . . . more than one member."[24] But consistent with that point we might say that Paula's existence in w3 falls within the neutral range *as does her existence at various other outcomes in various other cases* but that her existence in w2 falls below it.

---

[22] Broome 2004, p. 143.

[23] Broome, Stern Report contribution, p. 17.

[24] Broome 2004, p. 146.

The idea that Broome meant his argument to rule out the position that w3 and w1 are equally good might also seem at odds with his presentation of the inconsistency argument itself—its simplicity, its elegance, the seemingly obvious principles (transitivity, symmetry, the seemingly straightforward Pareto-like principle) that moved the argument forward. If the intuition Broome meant to prove inconsistent was *narrow* from the start, then the argument he would have needed to launch would have been considerably more complicated than the argument that he in fact describes.[25]

But consistent with that point perhaps Broome's statement of his own argument is itself just a sketch. Perhaps he takes for granted we'll fill in the gaps ourselves.

These two points together suggest we may have a *little* room to interpret his argument as targeting the claim that w3 and w1 are equally good—as, in effect, targeting what I am calling *narrow neutrality* here.

But there's still a third point in favor of that idea. It's *where Broome goes* once he's completed the inconsistency argument itself. Thus let's call the conclusion he reaches in the inconsistency argument—whatever the content of that conclusion—the *intermediate* conclusion. Broome then at various points seems to draw a further conclusion, an *ultimate* conclusion, to the effect that w3 is *better* than w1 is. That ultimate conclusion would indeed seem to follow—we'll see why in the next paragraph—*if* the intermediate conclusion itself is that it's not the case

---

[25] It would have been an *iterative* argument, one that would have involved the claim that just as w3 shows that Paula's existence at w2 isn't neutral, so does an outcome w4, where w4 is just like w3 except that w4 is better for Paula than w3 is, show that Paula's existence at w3 isn't neutral, and so on. And it would have been an argument that relies, not on the simple, straightforward Pareto-like principle that instructs that, in a case like the three option case, where w3 exists as an accessible outcome, w3 is better than w2, but rather a more contestable principle, one that asserts that w3 has a deflationary effect on the value of w2, making w2 worse than w1 even in the case where w3 *doesn't* exist as an accessible outcome to w2. We return to this question in part ___ below. But the upshot would be that the seemingly obvious proposition that w3 is better than w2 isn't really obvious at all if we stipulate from the start, not that w2 and w3 are accessible outcomes within the same case, but rather that the w2 we are talking about may hale from a different case altogether, one in which w3 *does not exist* as an accessible alternative.

that w3 and w1 are equally good—if, that is, the intermediate conclusion itself is just that *narrow* neutrality is false. But if all Broome has to work with there is that the neutral *range* claim is false, then that *ultimate* conclusion *doesn't* follow at all. It remains pie in the sky.

How would that intermediate conclusion, that w3 and w1 aren't equally good, help Broome get to his ultimate conclusion, that w3 is better than w1? Well, we really don't think that Paula's existence at w3 makes w3 *worse* than w1 is (we are not, after all, Benatarians). So let's—for the moment—take it as an assumption that that's so.[26] But if it's not the case w1 and w3 are equally good and it's not the case that w3 is worse than w1, then we are left to conclude that w3 must, after all, be *better* than w1 is—that Paula's existence in w3 must, after all, make things morally *better*.[27] In short: we should *agree* that, if it's not the case that w3 and w1 are equally good, then w3 is better than w1 is.

But that's a very strong ultimate conclusion, a conclusion with profound practical implications. Broome thus writes that "If [the neutrality intuition] were correct, it would give us a quick answer to the question about the value of extinction: it is neither good nor bad. *But actually the intuition is false*."[28] And since no one really thinks that, other things equal, the *non*-extinction of the species—the *survival* of the species—would make things *worse*—at least, so we shall assume for purposes here[29]—we are left to conclude that it would make things

---

[26] That w3 isn't worse than w1 really is just an assumption. It's entirely plausible given how we have here understood Broome's argument. However, on a reconstruction of that argument that we will consider later on, it's an assumption we shall need to question. See part ___ below.

[27] Again—contra Chang—I am taking for granted that in this simple case issues of comparability (or commensurability) do not arise. w1 and w3 are equally good, or w1 is worse than w3, or w3 is worse than w1. See note __ above.

[28] Toronto climate change remarks p. 8 (emphasis added). And "If the intuition of neutrality is correct, the extinction of humanity will be much less of a catastrophe than it might seem at first. . . . Actually, the intuition of neutrality has to be false. It cannot be consistently fitted into any theory of value." Broome, Stern Report contribution, p. 17.

[29] See note 28 above.

better. And: "*Given that the neutrality intuition is false*, the extinction of humanity might be a very great disaster indeed. It would prevent the existence of huge numbers of future people, and the existence of each one of them might well have been a good thing."[30]

Hence the question of interpretation. Is there more to the simple inconsistency argument itself than what we have so far seen? Should we understand it to purport to show, not just that it can't be that *both* w2 and w1 are equally good *and* that w3 and w1 are equally good, but rather that it can't be that w3 and w1 are equally good and hence, we agree, must be that w3 is better than w1? Is what seemed to be a simple inconsistency argument not really so simple after all?

Our purpose in life does not, of course, lie in interpreting Broome. Our purpose rather is to identify and then evaluate problems that might arise for *narrow neutrality*.

We accordingly face some worrisome possibilities. The first is that the simple argument isn't so simple after all—that it doesn't open the door to narrow neutrality and the position that w3 and w1 are equally good but rather annihilates narrow neutrality along with the position that w3 and w1 are equally good. The second is that we have gotten the inconsistency argument itself right but that there's a further argument that builds on that argument and that itself shows that it's not the case that w3 and w1 are equally good.

Whatever we find, the pressure to inquire can't be ignored. For it seems clear that Broome, somehow, thinks we can get to the result, not just that the neutral range claim is false, but that w3 is better than w1. It seems clear that on his view narrow neutrality is false. We need to understand just why that is so.

11.4   *Summing up.*   (i)   As conventionally formulated, traditional consequentialist theories—for example, *totalism*—imply that, other things equal,

---

[30] Broome, Stern Report contribution, p. 17 (emphasis added).

adding a person whose welfare level is positive makes a positive contribution to the total good of the world. We've noted totalism faces many problems.

(ii) The *neutrality* intuition comes along and claims that such contributions are *often* not positive but rather *neutral*. Broome's simple inconsistency argument—as original presented—effectively shows that that intuition—understood as the neutral *range* claim—is false. Consistent with that result, however, we can nonetheless accept *narrow* neutrality, w1 and w3 are equally good—that is, that Paula's existence in w3 is indeed neutral—but that w2 is worse than w1—that is, that Paula's existence in w2 makes things worse.

(iii)   We now face two alternate possibilities:

(iii.a) Per a further, not-yet-identified argument that we accept in place of the simple inconsistency argument or in addition to the simple inconsistency argument, we will be forced to reject that last claim—forced, that is, to reject not just the neutral *range* claim but also *narrow neutrality*; or

(iii.b) We won't identify any such further argument, or will identify *and* reject it, and thus be left with the room we need to *retain* narrow neutrality.

A full investigation of our options here requires us to understand a bit more about Broome's overall framework, including his construction of Harsanyi's principle, that is, P*. We turn to that work now.

Chapter 11

Additivity

11.1 *Broome's additive framework.* We can identify (at least) two further arguments that target *narrow* neutrality and aim to force the result that w3 and w1 aren't equally good and thus—on the assumption that w3 isn't *worse* than w1[31]— the result that w3 is *better* than w1. But a good understanding of how and whether those arguments work requires reference to aspects of Broome's work that go beyond his inconsistency argument.

We might, in any case, be interested in exploring Broome's overall framework for reasons that don't immediately relate to the procreative asymmetry or narrow neutrality. Additive in nature, Broome's overall framework comes with many of the plusses we earlier attributed to totalism. Yet by its very design it's meant to avoid some of the standard objections against totalism, including objections based on equality, fairness and, perhaps, priority.

Thus the first order of business in this Chapter 11 is to describe Broome's additive framework. Second, we identify two further arguments against narrow neutrality. And then third: we reject those arguments.

11.1 *Additivity, the personal good and the general good.* We've seen that Broome explores—and rejects—the correction to totalism proposed by the neutral range claim. But it isn't just our *existential* values that totalism—or indeed *any* view that calculates the total good of a world via a simple summation across individual welfare levels—seems impervious to. *Other* values that also seem left out of the picture by totalism but that, Broome concedes, a plausible theory may well need to recognize include values of *fairness*, *equality* and, perhaps, *priority*.

When Broome seeks a correction to totalism that is itself additive in nature, it's those *other* values, *not* our existential values, that he aims to show that additivity can accommodate. Let's see how the reconstruction works in connection

---

with those other values. We'll then ask the question whether that same reconstruction can be extended to cover our existential values as well.

Two steps are critical to Broome's reconstruction. The first is the concept of the *personal good*. The second is the connection between the *personal* good and the *general* good.

Thus Broome considers how an additive theory might accommodate the value of *equality*. Specifically, he considers how an additively *separable* theory might accommodate equality. Thus he doesn't, contrary to Temkin, see *inequality* as an impersonally defective *pattern* in a distribution of individual welfare levels at a particular outcome.[32] Rather, he sees inequality as something that is bad in a way that relates directly *to the individual* whose welfare level is lower at a given outcome when someone else's welfare level at that same outcome is higher. Both of those negatives—*both* the lower welfare *and* the inequality—might then be registered, according to Broome, in the *personal good* that we calculate for that person at that outcome.

In other words, *if* we think that inequality is morally significant, we have the option of understanding the *personal good* as reflecting, not just the bare fact that a person has a *lower welfare* level at a given outcome, but also the fact that that person can be considered a victim of a *failure of equality* (or of *fairness* or of *priority*) at that outcome.[33]

Thus the concept of the personal good is *highly* adorned and *highly* accommodating.

---

[32] Thus Temkin's hybrid, or pluralistic, theory might be *additive* in nature, but it's not *separately* additive: an impersonal pattern of inequality might detract from the value of a world, for Temkin, even if no person within that world can be counted in some sense a *victim* of that inequality.

[33] Fred Feldman has suggested a similar approach. Thus, in connection with the evaluation of outcomes, he proposes that the utility that an individual's existence contributes to the good of the outcome can itself be adjusted to take into account, e.g., justice. See *Pleasure and the Good Life*, pp. 195-197; and "Adjusting Utility for Justice: A Consequentialist Reply to the Objection from Justice," *Philosophy and Phenomenological Research* 55(3) (1995): 567-585.

The second step, then, is to make the connection between the *personal* good and the *general* good—that is, to use the concept of the personal good to transport the values of fairness, equality and priority into the additive picture. The value of equality is first captured in the concept of the *personal* good. It then exerts its influence on the evaluation of the outcome, or world, under scrutiny—the determination, that is, of the *general* good—via *summation*.

The upshot is what I will call P*.[34]

> P*. Where U(w) is the general utility of an outcome (or prospect) w and understood to represent the *general* betterness order between outcomes, where $u_1(w) \ldots u_n(w)$ are the personal utilities of the people in w, where a person's utilities "are defined to represent the person's betterness order" understanding betterness for the person as *personal* betterness, $U(w) = u_1(w) + u_2(w) + \ldots + u_n(w)$.[35]

To calculate the general utility of the world we simply add up the personal utilities that correspond to the personal good levels for each person who does or will exist in that world.

*   *   *

A theory can thus include the explicitly additive P* yet still recognize the value of equality since there's nothing in P* or any other component of Broome's framework that requires that the utilities to be summed up to determine the *general* good of a given outcome are the utilities that correspond to comparisons of people's *welfare* levels across a range of outcomes. Rather, the utilities to be summed up may be understood to correspond instead to a more complex comparison. Thus a person p's *welfare* may be the same in w1 as it is in w2 but w1 might still be *personally better for p* than w2, if, e.g., due to variations in the "conditions" of

---

[34] Here I closely follow Broome's own description of Harsanyi's principle. See note [12] above [Harsanyi's].

[35] Broome 2015, pp. 250-251.

other people in w1 and w2, p has as much welfare as other people have in w1 but less than they have in w2.

11.2 *Narrow neutrality, the personal good and inversion.* It's Broome's concept of the *personal good* and the connection he makes between the personal good and the general good that leads Broome to think that that P* does not rule out the values of equality, fairness or priority.

Does P* nonetheless rule out our *existential* values? Does it force us to say—on grounds entirely independent of the simple inconsistency argument and the assumption that w3 is surely not *worse* than w1—that w3 is *better* than w1? Does it rule out *narrow* neutrality?

Broome doesn't explicitly consider that question. I will just start by noting why we might think P* *doesn't* rule out our existential values.

Let's go back to the three outcome case. Per narrow neutrality, we say that Paula's existing at w3 does not make w3 *better* than w1 but that her existing at w2 does make w2 *worse* than w1.

The question now is whether Broome's additive framework rules out those happy results. Or, instead, can that same additive framework be understood to support those happy results? Can Broome's *highly* accommodating notion of the personal good be understood to reflect our *existential* values, just as, according to Broome, thinks it can be understood to reflect, e.g., our *egalitarian* values?

It seems, on the face of things, that it easily can. We simply take the position that Paula's *personal good* at w3 in point of fact falls at the single, sharp, neutral level despite the fact that her *welfare* level in w3 is positive. We can say, that is, that Paula's existence at the neutral level in w3 contributes *exactly* as much to w3's general good as Paula's never existing at all in w1 contributes to w1's general good—which is, of course, *none at all*. Summing up the relevant utilities—the utilities that correspond not to welfare but to the personal good—we then say that w3 and w1 are *equally generally good*.

To complete our account of the case, we take the position—indeed, must take the position—that, in addition, Paula's personal good at w2 falls in the *negative* range—again, despite the fact that her welfare level at w2 is itself positive. Summing the utilities now for w2, we say that w2 is *generally worse* than either w1 or w3.

In this way we can retain *both* narrow neutrality *and* P*. Of course, the account of the case we've just laid out commits us to a certain *inversion* in what might otherwise have seemed a natural way of understanding the personal good. It means that personal good levels will fall either at the *none at all* level or at the *negative* level. If the former, bringing the additional person into existence *doesn't* make the outcome *better* even if welfare is positive. If the latter, bringing the additional person, whether at a *negative* welfare level (as in the miserable child case) or at an *avoidably* low positive welfare level (as in the case of Paula at w2), may well make the outcome worse.

But there's no reason to think that inversion is problematic. Indeed, to declare inversion out of bounds from the start—without, that is, argument—would in effect beg the question against narrow neutrality. Inversion is the mechanism that allows us, within the additive framework, to retain narrow neutrality. Inversion *is* narrow neutrality. Moreover, it's quite sensible. A *negative* personal good level *doesn't* mean that the person's life isn't worth living. It just means that the world itself is *defective* in some morally significant way that is rooted in how an existing or future person at that world fares. And surely such a world *is* morally defective. Consider w2. More can be done there for Paula than has been done at no cost to anyone else at all; w2 thus plausibly is the morally lesser world and the wrong choice.[36]

---

[36] This Pareto principle needs to be spelled out very carefully. In the three outcome case, more is done for Paula at w3 than at w2, indicating a morally significant defect in w2. If we changed the case, and included a Quintus in w3 whose welfare in w3 is lower than it is in, say, some w4, then the condition on this simple Pareto principle would be failed: that would not be a case in which more can be done for Paula "at no cost to anyone else" since there would be a cost to Quintus in w3 notwithstanding the fact he never exists in w2.

Thus it seems on the face of things that we can readily understand Broome's framework and specifically his concept of the personal good, as accommodating, not just the values of equality, fairness and, perhaps, priority, but our existential values—that is, narrow neutrality—as well.

11.3 *Two arguments against narrow neutrality.* Broome's overall framework—including P*—now before us, we are in a position to try to identify further arguments that would help Broome target not just the neutral *range* claim but also *narrow* neutrality, and specifically, the claim that w3 and w1 are equally good. It might be tempting not to do that work and just to cheer narrow neutrality on. But we really can't comfortably *retain* narrow neutrality without first trying to identify and then evaluate Broome's arguments *against* narrow neutrality.

Broome's text suggests two such arguments. We'll start by briefly noting both arguments. We'll then examine each in more detail.

The first argument simply (i) looks at Paula's high welfare level in w3, (ii) notes that the position that her personal good level in w3 is *neutral* would mean that Paula's existence in w3 contributes just as much to the overall good of w3 as her never existing at all does and (iii) concludes that surely her personal good level in w3 must be at least a *little* greater than that!

Perhaps it's obvious that adding more detail to this first argument is not going to mean that it's not question-begging. We will consider what that detail might look like in what follows. But we may as well note now that the argument is likely *not* one Broome means to suggest.

The *second argument* may seem more promising. It starts with (i) the rejection of the neutral range claim. The argument thus takes as its first premise the conclusion of the simple inconsistency argument against the neutral range claim. So far so good. We then note that (ii) since the neutral level is, at most, a single, sharp, boundary, the odds are surely very much against anyone's ever coming into existence at exactly *that* level. Hence the odds are very much against Paula's coming into existence at exactly *that* level in w3. Put another way: the

odds are very much against Paula's existing at a *welfare* level of +10 in w3 itself means that her *personal good* at w3 contributes nothing at all to the general good of w3. The odds are indeed so low as to justify our ignoring the possibility altogether. The upshot? Paula *doesn't* exist at the single neutral level in w3. Given, then, the relation between the personal good and the general good—given, that is, P*—we conclude that w3 and w1 therefore *aren't*, after all, equally good.

But that second argument seems to fail as well. What makes +10 neutral *is the case*. In particular, it's that +10 is *maximizing* for Paula within the context of *the particular case*, that is, the *three outcome case*. The narrow neutralist thus would consider it no coincidence at all that the neutral level for that case and for that person would turn out to be +10. Anything less than that would be a *negative*; anything more than that would be *another case altogether*.

11.3.1 *First argument*. Let's take a closer look at the argument that I think Broome would *not* stand by. I did not, however, draw the argument out of thin air. There are some textual hints that favor it.

In the course of his discussion of the neutrality intuition, Broome underlines two closely related points. A person's existing at a *neutral level* at a given outcome means that that person's existing at that level contributes *exactly* as much to the general good of that outcome as that person's never existing at all at an outcome contributes to the general good of *that* outcome—that is, *none at all*. And, second, *any* level of the personal good—which is, just to underline, distinct from welfare; distinct, that is, from whatever it is that makes w3 better *for Paula* than w2 (or indeed w1) is—that is, in even the *slightest* degree, above the sharp boundary of the neutral level at a given outcome is such that the existence of a person at that level at that outcome will—by implication from P*—make that outcome generally better.

The text shows these points are central to Broome's discussion. Given the conclusion we are now after—that w3 and w1 are not equally good and that narrow neutrality is false—one might think the next natural step in the argument would

then be just this: in view of Paula's high welfare level in w3, surely her existence at w3 *must* come with a personal good level that exceeds *by at least some slight degree* the sharp boundary of the neutral level. Hence, by P*, adding Paula to w3 after all makes w3 better than w1.

But if that's indeed the argument, then the argument fails. Why should the narrow neutralist accept the claim that Paula's existence at w3 exceeds by *any degree at all*, slight or not, the single, sharp boundary of the neutral level? Certainly, that claim can't serve as an *assumption* of the argument. As an assumption, it's obviously question-begging. After all, the very issue we are trying to settle is whether Paula's existence at w3 makes w3 better, which itself, under P*, reduces to the issue of whether her personal good level at w3 exceeds the neutral level. Narrow neutrality claims that it doesn't.[37]

Perhaps, though, we can go still deeper and unearth a sub-argument for the otherwise question-begging claim that Paula's existence in w3 exceeds at least by a slight degree the neutral level. By hypothesis, Paula's *welfare* levels in w2 and in w3 are both positive. She is sufficiently well off in both w2 and w3 that the issue

---

[37] The argument, in other words, *assumes* that Paula's personal good in w3 itself is (at least very slightly) above the neutral level. But to *assume* or *stipulate* that Paula's personal good level in w3 is above the neutral level (or to claim it's positive on the grounds that her welfare level is positive or, indeed, let's suppose, at +10 *very* high) would be problematic (question begging) given the close (defined) relation between the personal good and the general good. After all, the question now on the table just is whether Paula's existence in w3 make w3 generally better than w1 is.

It is clear that Broome isn't aiming to foist off on us a question-begging argument favoring what seems to be his own clear conclusion on the neutrality intuition in general and extinction in particular: that is, that the neutrality intuition is false, and that it's not the case that extinction is neutral. At points, at least, he explicitly says that it's up to us to determine whether existence itself falls above or below or at the neutral level. ("Conceivably future people would on average life at the neutral level, in which case their existence together would be neutral. But that is such an unlikely coincidence we can ignore it. So the absence of all those future people will be either [personally and generally] good or bad. . . . I will leave this question unanswered." Broome, Toronto talk notes, p. 9.) Having been invited, then, to weigh in, narrow neutrality then does just that: despite her high welfare level in w3, Paula's personal good level in w3 is itself exactly neutral: w3 is neither generally better nor generally worse than w1 is.

of whether her existence has exactly the same value *to her* as her never existing at all would have had is settled; her existence, in both w2 and w3, from her own point of view is *well* worth having. Surely, then, her welfare level at w2, though lower than it is at w3, cannot be *so* low that her existence in w2 makes w2 generally *worse* than w1 is. Surely, in other words, w2 is *at least as good as* w1 even if *not better*. That would in turn mean that Paula's personal good level at w2 *cannot* itself fall into the negative range—that is, that the personal good her existence in w2 contributes to the general good of w2 via P* cannot be *less* that the personal good her never existing at all at w1 contributes to the general good of w1. But if that's the situation with her personal good level at w2, then—since it might seem that we can surely agree that her personal good level in w3 exceeds her personal good level in w2; the relation between welfare and the personal good may be complex but it's not *that* complex—we can infer that her personal good level at w3 after all *does* exceed the neutral level.

But that sub-argument fails. For—as noted earlier—to accept narrow neutrality is to accept *inversion*. It's just to accept that Paula's level of the personal good in w2 falls *below* the neutral level—falls, that is, into the negative range—*despite* the fact that her welfare level at w2 is unambiguously positive. Thus we may well agree that Paula's personal good level in w3 exceeds her personal good level in w2. But it's not going to follow that her personal good level in w3 exceeds the neutral level—that, in other words, her existence in w3 makes w3 better than w1 is.

Left without any adequate sub-argument for the claim that Paula's existence in w3 exceeds the neutral level, we should reject the first argument as question-begging. As noted before, it's doubtful that that first argument is one Broome meant to put forward to begin with.

11.3.2 *Second argument.* Let's now take a look at the second argument—the one I think Broome may well stand by.

The second argument begins with the conclusion of the simple inconsistency argument against the neutral range claim. The second argument thus starts with the point that the neutral level constitutes at most a single, sharp boundary—that is, that it can't be the case that *both* w2 and w1 are equally good and that w3 *and* w1 are equally good. We then simply note that any given person might exist at any one of perhaps infinitely many possible welfare levels. Paula's existing at +5 in w2 and at +10 in w3 are just two of those many levels. We then consider the odds against the proposition that her existence at +10 in w3 happens to coincide with *exactly that* single, sharp level. Surely they are very small—so small that we can safely "ignore" them altogether.[38] That in turn would mean that Paula's existence at w3 "will be either [generally] good or [generally] bad."[39] We've already accepted as an assumption that Paula's existence in w3 doesn't make w3 *worse* than w1 (we are not Benatarians).[40] We thus conclude that Paula's existence at w3 makes w3 generally *better* than w1 is.

Does this argument work? Can we on the basis of statistics dismiss the possibility that Paula's existence at w3 is itself neutral?

Let's step back. It seems clear that Broome's statistical argument does not even begin to look viable *unless* we eliminate the restriction that I included in my original presentation of the three outcome case and my original presentation of Broome's argument—that is, the *intra-case* restriction. According to that restriction, the three outcomes displayed in the three outcome case exist as *accessible* outcomes, and we are asked to compare w3 against w2 on the assumption that w3 isn't simply a remote *logically possible* world but rather an

---

[38] "Conceivably future people would on average life at the neutral level, in which case their existence together would be neutral. But that is such an unlikely coincidence *we can ignore it*. So the absence of all those future people will be either [generally] good or [generally] bad." Broome, Toronto talk notes, p. 9 (emphasis added).

[39] Broome, Toronto talk notes, p. 9.

[40] See note 28 above.

accessible world, a possible future agents had the ability, the power, the resources to make happen.

Moreover, the Pareto-like principle, articulated to include that same restriction, strikes us as simple, straightforward and indeed compelling. When w3 and w2 exist as *accessible* outcomes, we all agree that w3 is *morally better* than w2.

On the basis of that work, we then easily agreed that the neutral range claim is inconsistent.

We then pointed out that we could avoid the inconsistency by taking the position that w2 is worse than w1—and that we would then be free also to say that w3 is equally as good as w1. In other words: *within the particular case and for the particular person*, Paula, there exists at most a single, sharp neutral level of existence, that is, +10 at w3.

In fact, however, there is strong textual support[41] for the notion that the three outcomes Broome describes for purposes of constructing his inconsistency argument against the neutral range claim are *not* meant to be assumed to exist as accessible outcomes within the context of a given case—that is, that that argument is meant to proceed *without* any reliance on the intra-case restriction. The conclusion of the argument would then be considerably *stronger* conclusion—that, *for all outcomes whether accessible or not and (perhaps) for all people*, there exists at most a single, sharp neutral level of existence.

Adopting the intra-case restriction, we understand the argument to unfold within the confines of a particular hypothetical—a single case in which the issue is whether a given person is to be brought into existence at one of two distinct welfare levels or not brought into existence at all. Presented with that hypothetical, we are willing immediately to agree that w3 is better than w2 is.

---

[41] Broome himself states at one point that that is exactly the conclusion he means to reach. "But when we evaluate B in comparison to C, we must not assume B and C are actually available alternatives. Nothing says they are." Broome 2004, p. 147.

But then—as noted above—we're not going to consider against all odds that Paula's existence at +10 in w3 would happen to fall at the single, neutral level. What makes +10 neutral *is the case*. Paula's existence at +10 in w3 represents the *best* that can be done for Paula *in that case*; w3 is *maximizing* for Paula *in that case*. So of course *for that case* the neutral level will turn out to be +10, exactly the level at which Paula exists in w3.[42]

But now we are considering an alternate construction of Broome's argument. We are now considering the possibility that Broome means for us to *drop* the intra-case restriction—that he means his argument to reach for the conclusion that, for all people and all cases, there is a single, sharp, neutral level of existence.

That, in turn, would mean that, as we proceed to compare w3 against w2, for all we know, w3 hails from one case—the *three* outcome case—and w2 from different case altogether, a case in which w3 *does not exist* as an accessible outcome, that is, the *two*-outcome case.

---

[42] A point of clarification. It's true that *in one sense* there exists, for all people and all cases, a single neutral level: that is, that for all people and all cases, by definition, a person's existence at the neutral level in an outcome *by definition* contributes exactly as much personal good to that outcome as that the personal good a person's never existing at all in an outcome contributes to that outcome. The *neutral* level of the personal good is just the *none at all* level. But to think that that point itself means that there exists a single neutral level of *welfare* would be to confuse the *personal good* on the one hand and *welfare* on the other. That's a confusion that practically calls out to be made but it's a confusion all the same.

Two Outcome Case	Welfare	w1	w2
	+10		
	+5		**Paula**
	+0	*Paula**	

We are then asked to accept the claim that w3 in the three outcome case is better than w2 in the two outcome case.

Now, on this *inter*-case construction of Broome's argument, the point about just what a wild coincidence it would be for Paula's existence at +10 in w3 to fall at the single, sharp neutral level comes into play. For there will always be still another case—a case involving, e.g., an outcome w4 in which Paula's welfare level is greater than it is in w3, and a case involving outcomes w4 and w5 in which Paula's welfare level is greater in w5 than it is in w4 and greater in w4 than it is in w3. And so on. Given that potentially endless array of levels of welfare at which Paula *might* come into existence, why should we think that there's any real chance at all that Paula's existence at +10 in w3 would happen to fall at the single, sharp, neutral level?

Dropping the intra-case restriction thus may make Broome's statistical argument begin to look potentially viable.

In fact, however, there are still difficulties. We can reject the intra-case restriction and accept that the chances are very much against Paula's existence at +10 in w3 falling just at the neutral level. But the moment we reject the intra-case restriction the argument becomes vulnerable at another point. *Intra*-case restriction in place, the argument against the neutral range claim can proceed on the basis of a Pareto-like principle that is, due to the restriction itself, simple, straightforward and

compelling. When we understand that the case includes as accessible outcomes both w2 and w3, we are happy to say that w3 is morally better than w2 is. Without that restriction in place, the Pareto-like principle the argument asks us to put to work becomes much stronger. We are asked immediately to draw the inference that, even if w3 *doesn't* exist as an accessible alternative to w2, w3 is still *better*— *morally* better—than w2 is.

Now, on the face of things, that result may not seem objectionable. In fact, however, it's a result that is anathema to narrow neutrality. For given narrow neutrality's prior commitment to the position that w1 and w3 are equally good, this new result would then commit narrow neutrality to the position that w2 is *worse* than w1 in the *two* outcome case. But that's just not a plausible position. (Again, we are not Benatarians.) The reason adding Paula to w3 in the three outcome case doesn't, according to narrow neutrality, make w3 worse, or better, than w1 is just that Paula's welfare level in w3 has itself been maximized. That same reasoning applies to w2 in the two outcome case.

We shall thus want to say—though in a more exacting vocabulary[43]—that w2 being worse than w1 in the three outcome case does not imply that w2 is worse than w1 in the two outcome case. We shall, in other words, want to reject the stronger, inter-case version of the Pareto-like principle.

This point can be made entirely without reference to whether our Pareto-like principle is to be understood to be limited to the case where the outcomes we are ranking, w2 and w3, hail from the same case or from two different cases altogether. For purposes of developing the inconsistency argument against the neutral range claim, we are willing, whether on Pareto-like grounds or on other grounds entirely, to accept that w3 is better than w2. (Our maximizing intuitions are at play when we do that—but we can certainly get to that result without thinking that what is to be maximized is welfare in the *aggregate*.) But now let's make it explicit that the w2 we are asked to compare against w3 doesn't have w3 as an

---

[43] See part ___ below.

accessible outcome. There is, in other words, no w3 to exert a deflationary effect on w2; we have no grounds for saying w2 isn't equally as good as w1. There is no w3 such that we can say that w3 shows that w2 is worse than w1 is. On those facts, we are no longer willing to accept that w3 is better than w2 is.

There is still another problem with Broome's argument. We earlier conceded that surely w3 wasn't worse than w1. But if we accept the unrestricted Pareto-like principle, that concession shall need to be clawed back. If how w3 compares against w1 is to be determined, not by reference to the outcomes that exist as accessible to w3, but by reference to *all possible outcomes*, then given that, for any particular welfare level Paula has in any particular world, there is some possible world such that Paula's welfare level is at least a little higher, narrow neutrality would imply that all such worlds are actually *worse* than Paula's never existing at all. Broome might argue that that result shows that narrow neutrality cannot itself be correct. But we can't validly infer that result. For it's just as plausible—and indeed the position of the narrow neutrality—that how w3 compares against w1 is to be determined, not by reference to all *possible* outcomes, but rather by reference to all *accessible* outcomes—all outcomes, that is, that exist as accessible outcomes within the context of the particular case.

11.4 *Looking ahead*. What we are in effect saying here is that, in the three outcome case, w3 is better than w1 but that, in the two outcome case, w2 and w1 are equally good. But that way of looking at our facts may itself seem highly objectionable. It may seem to *over*-contextualize the discussion; it may seem to force us to the result that whether w2 is just as good as, or worse than, w1 can *vary* depending on the case, a result that in turn raises a host of theoretical issues. Thus we will need to make our substantive point in a considerably more exacting vocabulary. In the next chapter, we will thus attend to that issue and along with a handful of others. Assuming those issues can successfully be addressed, however, we will then be in a position to conclude that Broome provides us with a compelling

argument against the neutral *range* claim but no effective argument at all against *narrow* neutrality.

Chapter 12

Objections and Replies

12.1 *First objection*. If Paula's personal good in w2 is negative in the *three* outcome case, it must be negative in w2 in the *two* outcome case as well. That would mean, in turn, that it, after all, makes things *generally worse* to bring Paula into a perfectly fine existence in the two outcome case. However, while it's plausible to think that w2 is generally worse than w1 in the *three* outcome case, it's not at all plausible that w2 is generally worse than w1 in the *two* outcome case.

Reply: Narrow neutrality rejects the claim that, if Paula's personal good in w2 is negative in the *three* outcome case, it's negative in w2 in the *two* outcome case. Her *welfare* stays *constant* from one outcome to the other but consistent with that point her *personal good* in w2 in the two outcome case may be *greater* than it is in w2 in the three outcome case. That is, her personal good in w2 in the *two* outcome case may be exactly the same as her *personal good* level in w3 in the *three* outcome case (her welfare level having been maximized both in w2 in the two outcome case and in w3 in the three outcome case; there being nothing in the two outcome case to exert the same deflationary effect on her personal good level in w2 that w3 exerts in the three outcome case).

In taking the position that Paula's personal good in w2 in the two outcome case is the same as her personal good in w3 in the three outcome case—and that, correspondingly, the general good of w2 in the two outcome case is the same as the general good of w3 in the three outcome case—we remain in compliance with the rule that "[t]he value of a distribution depends only on the condition of each person; that is a consequence of the principle of personal good . . . . If the presence or absence of alternatives affects the value of a distribution, it can do so only by affecting some person's condition."[44] Thus I am not proposing that we calculate the personal good and then, depending on what outcomes are accessible, *reduce* the value of w2 in the three outcome case. Rather, I am proposing that in the three

---

[44] Broome 2004, p. 147.

outcome case we build the deflationary effect of w3 into Paula's personal good at w2 . . . and *then* determine the value of the outcome w2. In the two outcome case, there is no such deflationary effect on w2. So there is no basis on which to say w2 in that case is generally worse than w3 in the three outcome case.

We are here, in effect, *contextualizing* the value of Paula's existence in w2 and—with that, given P*—the evaluation of w2 itself. That is: the claim is that we can't fully assess the value of Paula's existence in w2, or determine whether w2 is worse than w1 or equally as good as w1, until we know what case we are in— until, that is, we know whether or not w3 exists as a further accessible outcome.

Broome himself notes that we can say that Paula is "wronged"—that is, her welfare level is avoidably reduced in a case where increasing her welfare could have been achieved at no cost to anyone else—in w2 in the three outcome case. That fact may provide us with grounds, as in the case of an inequality, for "reducing the value" of w2.[45] But we have no such grounds in the two outcome case. Hence we shouldn't, in that latter case, consider the value—that is, the general good—of w2 "reduced."

11.2 *Second objection*. w2 in the two outcome case *can't* be generally better than w2 in the three outcome case. Nothing relating to Paula's existence or the existence of anyone else has changed from one outcome to the other.

Reply: It's true w2 in the two cases is the same *for Paula from Paula's own point of view*—that is, that Paula's *welfare* level in w2 is the same in the three outcome case as it is in the two outcome case. But just as *welfare* can be the same for a subject in a given outcome from one case to another but due, e.g., to an unfairness, the outcome in the one case can be *personally better* for the subject than in the other case, *welfare* can be the same for Paula in w2 in both cases but w2 in the two outcome case can still be *personally better* for Paula than w2 in the three outcome case.

---

[45] Broome 2004, p. 147.

Now, this reply itself opens to the door to still another objection, an inconsistency objection. If Paula has a certain amount of personal good in w2, how can have more personal good than that in w2? How can she have more personal good in w2 in the two outcome case than she has in w2 in the three outcome case? We'll set this objection aside for now and return to it in part 12. 4 below.

11.3 *Third objection*. The dire facts I have built into the three outcome case are at odds with the ordinary case Broome has in mind when he refutes the neutral range claim. As I have constructed that case, Paula's personal good in w3 is "none at all" and in w2 it's actually negative. Broome, though, would have mentioned it if he had meant the case to include such dire facts!

Reply: Broome can't mean for us to *stipulate* as part of an argument against the neutral range claim that Paula's *personal good* in w2 and in w3, or even just in w3, is positive. Given the relationship between the *personal* good and the *general* good, such a stipulation would be problematic. After all, our question just is: does adding Paula at w3 (or w2) make that outcome generally better? Making it part of the original set up of the case that Paula's personal good in w3 (or w2) is *positive* would blatantly beg the question.

Moreover, once we distinguish welfare and the personal good, there's no basis for describing the facts I have built into the case as at all "dire." If characterizing her *personal good* as none at all at w3 and actually negative at w2 may seem a little glass-half-empty-ish—or as Johann Frick has put the point *harsh*—one can feel free to use the term *contributory value* in place of *personal good*.

11.4 *Fourth objection*. Narrow neutrality proposes what may seem to be a violation of the axiom of the independence of irrelevant alternatives, that is, the *independence axiom*. How can the mere *accessibility* of w3 leave w2 worse than w1 in the three outcome case but equally as good as w1 in the two outcome case?

In this connection, we also face a *consistency* question, one we deferred above: if Paula has a certain amount of personal good in w2, how can she at the same time have *more personal good than that* in w2? How can she have more personal good in w2 in the two outcome case than she has in w2 in the three outcome case?

Reply: Our discussion here can be brief. For this particular inconsistency argument is one that we have already seen. Thus, in the context of our discussion in Part I of whether any correct solution to the nonidentity problem was bound to abide by a certain axiological constraint, we considered whether the view that a comparison of one world w1 against a second world w2 in some cases depends on facts relating to still a third world w3 is consistent. There, as here, any such w3 that might affect, indeed, change how w1 compares against w2 will itself be an *accessible* world. But w3's accessibility relative to—say—w2 is a feature that can itself be discerned upon careful examination of w2. There is, we said, going to be a causal explanation of w3's accessibility—an explanation that is itself rooted in the *modal details* inherent in w2: how things, within the bounds established by, e.g., the laws of nature and (perhaps) the acts of other agents, *could* have been. By the same token, when w3 isn't accessible relative to w1, that, too, is going to have a causal explanation, itself rooted facts about w2. To say that w3 isn't accessible in that case is just to say that agents in w2 *lacked* some power, some ability, to make things any better for p, that is, to bring about an alternate possible future that includes the advantages for Paula we see in w3.

Having come this far, we can then easily see our way clear to the next step. That agents have the relevant ability in the one world and lack that ability in the other just means that those worlds—w2 in the three outcome case and w2 in the two outcome case—are actually two *distinct* worlds. Worlds, after all, aren't simply *distributions*—bare boned assignments of welfare levels to members of a particular population. Rather, worlds come to us with all their details necessarily intact. New details entail new worlds.

A more exacting vocabulary will recognize exactly that point—and the inconsistency we were worried about never in fact arrives. Thus we might say that Paula indeed has less personal good in w2 than she has in w1 in the three outcome case due to the accessibility in that case of w3, which accessibility is itself reflected in the modal details inherent in w2 and w1. But she has exactly the same amount of personal good in w2′ than she has in w1′ in the two outcome case—and this last, despite the fact that w2′ and w2 distribute welfare across exactly the same population in exactly the same way.

What secures this result—guaranties, that is, that we won't come across still another case a case in which we seem bound to recognize an identity between w2 and w2′ and hence face inconsistency all over again—is the *accessibility axiom*.

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

Avoiding inconsistency by introducing a more exacting vocabulary means that we can retain the independence axiom understood in a certain way. If the principle means we aren't allowed to look closely enough *at w1 and w2* to see whether *w3* is indeed accessible—if it means we must blind ourselves to those particular facts about w1 and w2—then independence must go. But if it's understood as imposing not *quite* such a ridiculously strict standard as that, then it may stay.

Chapter 13

Conclusions; Implications

13.1 *Inversion and narrow neutrality.* A main purpose here has been to suggest that we might save the only version of the neutrality intuition that we want to save—that is, *narrow neutrality*—through an *inversion* of the picture we perhaps first imagined when presented with the neutrality intuition. Rather than thinking of differences in welfare levels as having some effect on general good in some range *above* the neutral level, we can think of those differences as having some effect on the general good in some range *below* the neutral level.

The tension all along for the neutralist—that is, for the theorist who wants via some form of the neutrality intuition to take existential values into account; that is, for the *existentialist*—is to explain why Paula's higher welfare level in w3 has a critical impact when we compare w3 against w2 but no impact at all when we compare w3 against w1. Why is it *something* when we compare w3 against w2 and *nothing* when we compare w3 against w1? Why is her higher welfare level so potent when it comes to the one comparison but so completely ineffectual when it comes to the other? How can a mere shift in the *question we happen to be asking* change the *value* of Paula's higher welfare level? How can Paula's higher welfare level in w3 make w3 better than w2 *without* making w3 better than w1?

The *inversion* that the neutrality intuition puts into place resolves that tension. It does so by setting the personal good level for Paula's existence at w3 at the *none at all* level—at, that is, the level that is exactly the same as the level of personal good Paula's never existing at w1 contributes to the general good of w1— and setting the personal good level for Paula's existence at w2 *below* the neutral level. That means that the personal good Paula has at w3 can stay perfectly constant at the *none at all* level whether we happen to be comparing w3 against w2 or happen to be comparing w3 against w1. And that in turn means we can say—in fact it *requires* us to say given P*—that her existing at just that level in w3 indeed makes w3 better than w2 but does not make w3 better than w1.

A related tension arises in connection with our evaluation of w2. Why is Paula's existence in w2 perfectly innocuous when we compare w2 against w1 but morally troubling when we compare w2 against w3? Larry Temkin attempts to resolve that tension by bringing into the analysis of the three outcome case two very distinct sorts of views, the *Internal Aspects View* and the *Essentially Comparative View*. The former suggests that w2 is at least as good as w1 is (based on the internal aspects of each of the two outcomes and without reference to what is going on in any third outcome) while the latter suggests that w2 *isn't* at least as good as w1 is (based on an examination of outcomes *beyond* w1 and w2; based, that is, on the accessibility of w3).

I see narrow neutrality as supportive of what Temkin is aiming to accomplish in taking the position that the evaluation of the three outcome case implicates both the Internal Aspects View and the Essential Comparative View. In contrast, though, to Temkin's approach, which ultimately involves weighing a plurality of values against each other, narrow neutrality, by setting (in the three outcome case) Paula's existence in w2 below the single neutral level whether we are attending to w3 as an available alternative outcome or not, will exclude the problem result: that is, the result that w2 is at least as good as w1 is. (Temkin himself, I should note, may view that point as a deficiency in, not an advantage of, narrow neutrality.)

13.2 *The nonidentity problem.* In Part I, I argued that the nonidentity problem did not present, on further examination, a problem for the person-affecting intuition, that is, the intuition that a "bad" act must be "bad for" some person or another who does or will exist. There was a proviso on that conclusion. For it to hold, the person-affecting intuition had to be understood in a certain way. PAIA(c) wouldn't do; we needed PAIA* instead.

It's interesting now to see how my proposed solution to the nonidentity problem, which insisted that we take into account, in determining whether a given

act is "bad for" a given person or not, the full modal array, and not some arbitrarily limited subset, can be expressed under a theory of narrow neutrality.

Thus we can say that Andy and Rachel exist at the neutral level in w2 and w3, respectively, but that Andy exists below the neutral level in w1. Thus the personal good contributed by Andy's and Rachel's existence in w2 and w3 to the general good of w2 and w3 is at the "none at all" level, whereas Andy's personal good in w1 actually falls into the negative range, meaning that his existence there actually takes away from the overall good of w3.

We can use neutral colors – grays, whites etc. – to represent personal good at the "none at all" level and potentially deepening shades of red to represent negative personal good. Thus:

Graph __: Narrow Neutrality and the Modally Enriched Presentation of the Nonidentity Problem			
welfare	w1 (including a1)	w2 (including a2)	w3 (including a3)
+10		**Ruth**	**Andy**
+8	**Andy**		
+0	*Ruth**	*Andy**	*Ruth**

13.3 *Infinite population problem.* If we overlook the possibility of inversion we limit our capacity for cogent analysis. By restricting our answers to how much Paula's situation adds to the overall good of each outcome to the *positive*—if we think of it as positive in w3, positive, though lower, in w2 and *none at all* for w1—we handcuff ourselves. If we instead recognize the contribution the

addition of Paula makes to w2 as *below* the neutral level, we become able to say sensible things about that case.

The inversion we obtain from narrow neutrality doesn't just help us in the three outcome case. There are other cases as well in which that strategy facilitates analysis. Consider, e.g., the *infinite population problem*. In that problem, we are to imagine an infinite population existing at a relatively low, though still clearly positive, welfare level in w1, and that same population existing at a significantly higher welfare level in w2. Totalism, of course, immediately seems defeated by this case, since it counts w1 and w2 as equally good, whereas it seems intuitively clear that w2 is morally better than w1. That's especially so, if we do imagine how things look from the perspective of the individual members of the population.

More generally, the infinite population problem is difficult if we think of each of the infinite lives worth living as adding something *positive* to the overall good of the relevant outcome. We are tempted to reconstruct the case in a way that has us apply the additive principle not to the infinite set but rather to selected finite subsets of that infinite set. But if, for the outcome w1 in which each person's welfare level is, say, +1, we think of those person's lives as adding something *negative* to w1—if, that is, we think of the personal good level as falling into the *negative* range—and, for the outcome w2 in which each person's welfare level is +2, we think of those person's lives as adding *nothing at all* to w2, we then have a basis for the claim that w2 is generally better, indeed *infinitely* better, than w1 is— which is, of course, exactly what we would like to say about that case.

Again using grays to represent welfare at the neutral range, and reds to represent welfare that falls below the neutral range, we can sum up as follows:

Infinite Population Problem	Welfare	w1	w2
	+2		**p1, p2, p3….**
	+1	**p1, p2, p3….**	
	+0		

13.4 *Replaceability*. What about the poor little dog we thrust into the limelight in the introduction to this book and have since lost track of completely? Let Dolly be the little dog. The question is to compare an outcome in which Dolly's life is saved how the outcome in which Dolly's life is saved compares against the outcome in which Dolly is painlessly euthanized and a distinct little dog Jolly is created in the lab that is so similar to the original that—like a new pet minnow— the family itself will never know the difference.

I am taking for granted here that we reject the totalist account of this case. What we want to know, rather, is whether we can make sense of our intuition that w2 is actually better than w3 notwithstanding our commitment both to maximization and to impartiality. My view is that we can. While Dolly's welfare in w2 is identical to Jolly's in w3, it's also the case that Jolly's existence at w3 can plausibly be viewed, under narrow neutrality, as contributing nothing at all to the general good of w3—that is, that her personal good at w3 is +0. At the same time, under inversion, we shall want to recognize that Dolly's existence at w2 itself detracts from the general good—that her personal good level at w2 actually falls into the negative range.   No one else being affected either way, we may immediately conclude that w2 is generally better than w3 is. What of w1? We are free to say there that Dolly's personal good level is, as it is in w3, in the negative range, and hence, again simply adding things up, that w1 is actually worse than either w2 or w3. Accepting the very close connection between betterness and what

we ought to do, we conclude, finally, that the act that results in w2 is obligatory and the other two just wrong.

Summing up:

Replaceability	w1 (we do nothing)	w2 (we cure Dolly)	w3 (we painlessly euthanize Dolly and create Jolly)
+8		**Dolly**	**Jolly**
+4			**Dolly**
+0	**Dolly**, *Jolly**	*Jolly**	

Accounts of other problem cases in population ethics are explored in Appendix __ below.

Key: Bold face means the indicated person does or will exist in the indicated outcome; italics/* means indicated person never exists in indicated outcome.

Implications of *narrow neutrality* (in combination with other plausible principles) for levels of the personal good are shown by color. A white surface area indicates that the person's personal good level is neutral. Potentially deepening shades of red indicate that the person's personal good level is *less* than the neutral level—that is, that (given P*) that person's plight (or existence) in that outcome *takes away* from the general good.

Basic case	Welfare	w1	w2
Life well worth living	+10	**George**	**Jill**
	+9 … +2		
Life barely worth living	+1		**George**
	+0	*Jill**	

Graph ___: Addition Plus			
Welfare	w1	w2	w3
+11		**p**	
+10	**p**		
+5			**p, q**
+1		**q**	
+0	*q**		

Assumption: In Addition Plus, priority view allows us to say that that p's existing in w3 (at welfare +5) takes less away from the overall good than q's existing in w2 (at welfare +1) does.

Graph ___: Mixed Existence			
Welfare	w1	w2	w3
+4		**q**	**P**
+3		**p**	**q**
+0	*p*, q**		

Puzzle: In Mixed Existence, we want to say w2 = w3 = w1. So can't assign a personal good level below the neutral level to p in w2 or to q in w3. The puzzle then is why it's consistent for us then to assign a personal good level below the neutral level to q in w2 in Addition Plus. The basis for that "negative" assignment in Addition Plus is that q exists in w2 and has less welfare in w2 than q has in w3. But that basis exists in Mixed Existence as well. Solution: there is a difference in the two cases. In Mixed Existence, the only way to make p better off than p is in w2 is to make q exactly as badly off as p is in w2. In Addition Plus, there is a way of making q better off than q is in w2 that does not make anyone as badly off as q is in w2.

Double Wrongful Life (Parfit)	Welfare	w1	w2
	+0	*p**	*q**
	-10	**q**	**p**

Multiple Wrongful Life	Welfare	w1	w2	w3
	+0	*p*, q**	*p**	
	-10	r	q, r	p, q, r

Note on multiple wrongful life: Retention of P* over the personal good means that if we multiply wrongful lives we will be making each successive outcome worse than the prior outcome.

Single Wrongful Life	Welfare	w1	w2
	+0	*p**	
	-10		**p**

Personal good of p in w2 in Single Wrongful Life is distinct from personal good of p in w2 in Double Wrongful Life even though welfare levels in the two outcomes in the two cases are identical. Why? Because personal good is determined by factors beyond simple welfare, including conditions of other people's existence (e.g. conditions of q's existence in w1 in Double Wrongful Life).

Tom, Dick and Harry (Parfit)	Welfare	w1	w2	w3
	+4	**Harry**	**Dick**	**Tom**
	+2	**Dick**	**Tom**	**Harry**
	+0	*Tom**	*Harry**	*Dick**

For same reason we conclude in Mixed Existence that we wouldn't assign personal good below the neutral level to p's existing in w2 or q's existing in w3, we want to say here as well that we won't assign personal good below the neutral level to Tom's existing in w2 (Harry's existing in w3, etc.).

APPENDIX B

First Argument: Broome's Argument Against the Neutral Range Claim

Let "=" between outcomes mean that each outcome *is equally as generally good as* the other; let "<" between outcomes mean that the first outcome *is generally worse than* the second; and let "<" otherwise have its usual meaning.

Consider a case in which a person Paula never exists in w1; exists and has m level of welfare in w2; and exists and has n level of welfare in w3. Assume that m<n and that w1, w2 and w3 are otherwise the same (same population; same distribution). Assume that w3 exists as an accessible outcome relative to w1 and w2. The question is whether Paula's existing at welfare level m in w2 and Paula's existing at welfare level n in w3 can *both* count as Paula's existing at the neutral level, i.e., as instances in which her existence does not make the outcome itself better (or worse).

1. In this single case, there exist at least two neutral levels m and n.	Assumption of neutral range claim for *reductio*; facts of case
2. w1 = w2.	(1), definition "neutral level"
3. w1 = w3.	(1), definition "neutral level"
4. w2 = w3.	(2), (3), transitivity and symmetry of =
5. w2 < w3.	Pareto-like principle (restricted to the case where w3 is accessible to w1 and w3)
6. Inconsistency	(4), (5)
7. The neutral range claim is false; m and n can't both be neutral levels.	Reductio (1)-(6)

Lines (1)-(4) seem unobjectionable. But a note on line (5) is in order. Given that m and n designate levels of welfare, not the personal good, and given the intra-case restriction— that we are in a case in which w3 exists as an accessible outcome relative to w1 and w2— (5) seems unproblematic. The upshot is that the argument goes through—that the neutral range claim (intra-case restriction in place) is false.

Broome justifies line (5) by reference to the principle of personal good (PPG). Despite its title, there really is a question whether PPG is intended to talk about simple *welfare* levels or levels of the *personal good*. Let's consider the first reading first. That reading makes PPG equivalent to the Pareto-like principle I have cited as justifying line (5).

The principle of personal good (*PPG*). Where two outcomes have the same population, if one outcome assigns at least as much welfare to each member of the population as the other is and more welfare to at least some member of the population than the other does, then the one outcome is generally better than the other (WL p. 120, rewriting using GPG vocabulary).

Given PPG, so interpreted *and* taken together with the restriction I have included in the justification column for (5), line (5) seems unobjectionable. After all, PPG is *both* a "same-population" *and* a "same-person" principle. It's explicitly limited to the case where two outcomes share exactly the same population. And, again explicitly, the sufficient condition is satisfied only if the first outcome is personally better for a person than the second outcome is *for that same person*.

Given the inconsistency in (6), Broome rejects the neutral *range* claim in favor of the *single neutral value claim*:

Single neutral value claim: There exists at most one neutral level of the personal good—one level of the personal good such that adding a person at that level makes an outcome neither generally better nor generally worse.

That is: for *any one person* within *any one case* that includes the details we have included here—same population; better for at least one and worse for none; no outcomes beyond the three outcomes described—there exists at most a single neutral level. With those caveats in mind, I am happy to accept the argument and the conclusion.

Let's now consider the second reading. To generate the second reading, we simply substitute personal good in for welfare throughout PPG.

But now we have an even quicker argument. If m and n talk about the personal good, and m < n, then just in virtue of how P* defines the terms it follows that w2 < w3. It can't be, then, that both m and n are neutral levels. So again the neutral range claim is defeated. But notably the account would leave two options open: either m falls into the negative range, or n falls into the positive range. Narrow neutrality would favor the former option. In any case, assuming the second of these two options holds would be to beg the question against narrow neutrality.

Second Argument: Argument Against Narrow Neutrality

There is only one neutral level.	Single neutral value claim
Adding Paula at the neutral level adds exactly as much personal good to an outcome as Paula's never existing at all adds to an outcome.	Definition "neutral level"
Adding Paula at any level of the personal good *greater* than the level of personal good that Paula's never existing at all adds makes an outcome generally *better*.	1, 2, P* (more personal good entails more general good)
Paula's personal good at level n is greater than the level of personal good that Paula's never existing at all adds to an outcome.	Facts of case (Paula's welfare level in w3 is +10; she has a very good life; not even close to the sharp boundary below which lives aren't worth living)
Adding Paula to w3 makes w3 generally better.	4, relation between personal good and general good, P*
Personal good levels "often"—within limits—are greater than the level of personal good that a person's never existing at all adds to a given outcome.	5, universal generalization (nothing special about this case)
The neutrality intuition is false; "often"—within limits— adding a person makes the world generally better.	6

This argument I take not to be Broome's argument (though he might, recognizing (4) as a mere assumption, accept the *conditional* that, if (4) holds, adding Paula to w3 in the three outcome case makes w3 generally better). As the argument stands, however, we may reject the conclusion on the grounds that we have no basis on which to accept (4) and any stipulation that (4) holds would be question-begging against the view that adding Paula to w3 doesn't make w3 generally better. [END]