

Person-Affecting Views and Saturating Counterpart Relations

Christopher J. G. Meacham

Forthcoming in *Philosophical Studies*

Abstract

In *Reasons and Persons*, Parfit (1984) posed a challenge: provide a satisfying normative account that solves the Non-Identity Problem, avoids the Repugnant and Absurd Conclusions, and solves the Mere-Addition Paradox. In response, some have suggested that we look toward person-affecting views of morality for a solution. But the person-affecting views that have been offered so far have been unable to satisfy Parfit's four requirements, and these views have been subject to a number of independent complaints. This paper describes a person-affecting account which meets Parfit's challenge. The account satisfies Parfit's four requirements, and avoids many of the criticisms that have been raised against person-affecting views.

1 Introduction

Interesting ethical questions arise when we consider decisions that bear on the makeup of the overall population, present and future. Traditional moral theories tend to yield highly counterintuitive results when applied to these kinds of cases, and finding alternatives to the traditional theories that avoid these counterintuitive results is surprisingly difficult. The resulting state of affairs is nicely described by Derek Parfit, who ends his exhaustive examination of these issues with the following summary of his investigations:

“We need a new theory of beneficence. This must solve the Non-Identity Problem, avoid the Repugnant and Absurd Conclusions, and solve the Mere-Addition Paradox. I failed to find a theory that can meet these four requirements.”¹

Although these remarks concern his own inquiries, they could reasonably be said to represent the prevailing opinion regarding this literature as a whole.² In response to these

¹Parfit (1984), p.443.

²Of course, many have denied that all of these requirements *need* to be met, and have gone on to endorse theories which satisfy some subset of these requirements. See Ryberg, Tännsjö and Arrhenius (2009) for a comprehensive discussion of the many different responses that have been offered to Parfit's dilemma.

problems, some have suggested that we look toward “person-affecting” views of morality for a solution—views whose evaluations are sensitive to the identities of the subjects in the different possible outcomes.³ In order to provide a satisfying person-affecting treatment of these issues, two things need to be done.

First, one needs to address the many criticisms of person-affecting views that have been offered in the literature. These kinds of theories have been argued to be either inconsistent, highly counterintuitive, or unhelpful with respect to the original problems.⁴ More generally, critics have maintained that person-affecting views are unable to satisfy all of Parfit’s requirements. And a satisfying person-affecting response to the issues Parfit raises must either rebut these criticisms, or provide an account that avoids them.

Second, one needs to determine how to identify subjects in different possibilities. This is because the prescriptions of person-affecting views depend crucially on how we cross-identify subjects in different possibilities.

These two tasks are not independent. Certain approaches to the second task will require us to re-evaluate whether the standard criticisms of person-affecting views still arise. In light of this, it’s natural to wonder whether there is a way of tackling both tasks at the same time. I.e., one might adopt an account of how to identify subjects in different outcomes that allows one to circumvent the criticisms that have been raised against person-affecting views.

This is precisely what I propose to do. I will sketch a person-affecting view, the *Harm Minimizing View*. Then I will sketch a way of pairing subjects in different possibilities using what I call *saturating counterpart relations*. I will suggest that we should use these kinds of counterpart relations when making person-affecting judgments. We can then combine this person-affecting view with this way of pairing subjects. The resulting combination yields a person-affecting approach that satisfies Parfit’s four requirements, and which avoids many of the criticisms that have been raised against person-affecting views.

The rest of this paper will proceed as follows. In the next section, I will briefly lay out some preliminary assumptions. In the third section, I will lay out the Harm Minimizing View. In the fourth section, I will turn to examine the Non-Identity Problem. While doing this, I’ll describe and motivate the adoption of saturating counterpart relations. In the fifth section I will examine the Repugnant Conclusion. In the sixth section I will examine the Absurd Conclusion. In the seventh section I will examine the Mere-Addition Paradox. While doing so, I’ll describe and assess a powerful decision-theoretic objection to approach I advocate. I’ll also show why the various “impossibility theorems”—theorems which show that no theory can satisfy all of some desirable set of features—do not tell against this approach.⁵ In the eighth section I assess some other potential objections. In the ninth section, I conclude with some brief remarks.

³For example, see Narveson (1967) and Roberts (1998).

⁴For example, see Parfit (1984), Broome (1992), Arrhenius (2003) and Holtug (2004).

⁵For example, see Ng (1989), Blackorby and Donaldson (1991), and Arrhenius (2000).

2 Preliminaries

I'll assume that we have established some account of who the moral patients are; i.e., of who matters morally. When I speak of individuals, subjects, etc. in the sections that follow, I will be implicitly restricting myself to such beings.

I'll assume that there is some sense in which moral patients can be “well-off” that is morally relevant. And I'll assume that there is some way of providing an overall lifelong assessment of how well-off these moral patients are; I will call this the patient's *well-being*.

I'll assume that the level of a subject's well-being can be given a numerical representation. And I'll assume that there is some level of well-being below which a life is not worth living. In what follows I'll employ a numerical representation for well-being which is additive, and whose zero-point is set so that positive values represent lives worth living and negative values represent lives not worth living.

I'll assume we can make sense of what an agent's *options* are at a time. For simplicity, I'll also assume that every option available to an agent leads to a definite outcome, and that the agent knows this. So I'll ignore any role that chance and uncertainty might play.

I'll skirt issues involving infinities by restricting my attention to finitary cases. In particular, I'll assume that (i) agents are faced with only finitely many options at any given time, (ii) there are only finitely many subjects in any given possibility, and (iii) the well-being of these subjects is finite.

Finally, I'll assume that something like Lewis' counterpart theory is correct.⁶ On counterpart theory, the truth values of *de re* modal claims are cashed out in terms of counterpart relations between possible individuals (*a* is a counterpart of *b*). For example, let “Bob” be the name of some possible individual. Then “Bob could have been a plumber” is true *iff* some counterpart of Bob is a plumber.⁷ Likewise, “Bob is essentially human” is true *iff* every counterpart of Bob is human.

On Lewis' theory, counterpart relations are similarity relations. A possible individual is a counterpart of another *iff* the intrinsic and extrinsic qualitative properties of the former resemble those of the latter in the relevant respects. The kinds of properties that are relevant, and the stringency of the resemblance that's required, is something that can vary from context to context.

Note that counterpart relations are generally not symmetric—*b* may be a counterpart of *a*, even though *a* is not a counterpart of *b*. Likewise, counterpart relations are generally not transitive—*b* may be a counterpart of *a* and *c* may be a counterpart of *b* without *c* being a counterpart of *a*.⁸

⁶See Lewis (1986). Although I'll be assuming that Lewis' theory is correct in broad outlines, I will not be assuming that he is right regarding all of the particulars; c.f. section 4.

⁷Or, more precisely, “Bob could have been a plumber” is true *iff* there is some world *W*, and some counterpart of Bob in *W*, which is a plumber (see Lewis (1986), p.9-10). Similar remarks apply to the example that follows.

⁸To see the former, note that *b* may be the individual at *b*'s world that most closely resembles *a*, but there may be other individuals at *a*'s world that more closely resemble *b* than *a* does. To see the latter, note that *b* may be similar enough to *a* to be its counterpart, and *c* may be similar enough to *b* to be its counterpart, but the resemblance gap between *c* and *a* may be too wide for *c* to be *a*'s counterpart.

3 The Harm Minimizing View

In this section, I will describe a person-affecting view, which I'll call the *Harm Minimizing View*. This view is similar to a number of other person-affecting views that have been described in the literature, such as those of Roberts (1998) and Arrhenius (2003).⁹ For exegetical purposes, I'll present the view in two stages. First I'll sketch the view for cases in which all of the outcomes contain the same individuals. Then I'll extend the account to cases in which outcomes have different individuals.

3.1 Same Population Cases

To begin, let's restrict our attention to cases where all of the potential outcomes contain the same individuals.

At first pass, we might characterize the person-affecting intuition as follows: in order for an option to be better or worse than another, it has to be better or worse *for* someone.¹⁰ So suppose an agent is choosing between two outcomes, W_1 and W_2 . When we compare the W_1 -option to the W_2 -option, we should consider, for each subject, how much better or worse-off she is in W_1 than in W_2 .

To turn this into a concrete proposal, we need to determine which subjects are better-off in which outcomes, and to turn this into a judgment about what the best options are. Let's look at a way to do this.

Consider all of the outcomes that an agent a could bring about at a given time t . In some of these outcomes, a given subject s will have a higher well-being; in others, a lower well-being. Let's call the highest well-being that s receives in any of these outcomes s 's *peak well-being* (with respect to a at t). If s 's well-being in some outcome W_1 is below her peak well-being, then there's a sense in which bringing about W_1 harms s .

We can use this notion of harm to assess an agent's options. Let the *harm done by the W_1 -option* (with respect to a at t) be equal to the sum, for each of the subjects in W_1 , of the amount by which that subject's well-being is below her peak. Then we can evaluate an agent's options as follows:

The Harm Minimizing View (HMV): An option is morally permissible (for a at t) *iff* no other option does less harm; i.e., *iff* the option minimizes harm.

Example: Weighing Losses. Consider an agent who has a choice between two outcomes, W_1 and W_2 . In W_1 there will be two individuals, a and b , each with a well-being of +10. In W_2 there will be two individuals, a and b (where giving two individuals the same name indicates that each is a counterpart of the other). But in W_2 , a will have a well-being of +15, while b will have a well-being of 0. Visually, we can represent this case as follows:

⁹In the case of Roberts (1998), the similarity is less apparent. But one can think of the Harm Minimizing View as a quantitative version of Roberts' view. And the prescriptions of the two views are almost identical (though Roberts' view is silent in some cases in which the Harm Minimizing View is not).

¹⁰“At first pass” because satisfying this description is neither necessary nor sufficient for being a person-affecting view. For discussion regarding different ways of spelling out the person-affecting intuition, see Arrhenius (2003), Roberts (2003b) and Holtug (2004).

W_1		W_2	
a	b	a	b
+10	+10	+15	0

According to HMV, what should the agent do? In this case, a 's peak well-being is +15, while b 's peak well-being is +10. In W_1 , a 's well-being is 5 units below her peak, while b 's well-being is 0 units below her peak. So the harm done by the W_1 -option is: $5 + 0 = 5$. In W_2 , a 's well-being is 0 units below her peak, while b 's well-being is 10 units below her peak. So the harm done by the W_2 -option is: $0 + 10 = 10$. Since the W_1 -option does 5 units of harm and the W_2 -option does 10, HMV prescribes the W_1 -option.

3.2 Different Population Cases

Now let's look at cases in which different outcomes contain different individuals. Consider a choice between two outcomes, W_1 and W_2 , where some subject s comes to exist in W_1 but not W_2 . How should s 's existence bear on our assessments of these options?

There are two natural ways to proceed.

First, one might maintain that s 's existence should have no bearing on the harm of the W_1 -option, regardless of what s 's well-being happens to be. Since s only exists in W_1 , s 's peak well-being will be whatever s 's well-being in W_1 is. Thus s will have her peak well-being, and won't add to the harm done by the W_1 -option. (She also won't have any affect on the harm done by the W_2 -option. The harm done by the W_2 -option is an assessment of how much the subjects in W_2 are below their peak, and so will only take into account subjects who are in W_2 .)

What about the harm done by the W_2 -option? That's an assessment of how much the subjects in W_2 are below their peak, so that will only take into account subjects who are in W_2 . So s 's existence in W_1 won't have any affect on the harm done by the W_2 -option.

Second, one might maintain that s 's existence *can* have a bearing on the harm done by the W_1 -option. In particular, suppose that s 's well-being in W_1 is so low that s 's life isn't worth living. Then there's a sense in which s can claim to have been harmed if W_1 comes about. After all, the agent could have picked the W_2 -option, and s would not have existed. But the agent picked the W_1 -option instead, and now s is forced to live a life not worth living. (Again, s 's existence will have no bearing on the harm done by the W_2 -option, since this value only considers subjects who exist in W_2 .)

Some have argued that approaches like the second are incoherent.¹¹ But detailed and compelling responses to these arguments have been offered in the literature.¹² And the second approach fits better with person-affecting approaches. So I will adopt the second approach here.

We can implement the second approach by modifying the characterization of a subject's peak well-being given earlier. Let s 's *peak well-being* (for a at t) be the highest well-being that s receives in any of the available outcomes, where for these purposes, s is

¹¹For example, see Broome (1999) and Arrhenius (2003).

¹²For example, see Parsons (2002), Roberts (2003a) and Holtug (2004).

treated as having a well-being of 0 in outcomes where she doesn't exist. This modification will yield the verdicts we want.

Example: The Question of Creation. Consider an agent who has a choice between two outcomes—creating no one, or creating both a happy person and an unhappy person:¹³

W_1	W_2	
	a	b
	+5	-5

According to HMV, what should the agent do? For the purposes of determining peak well-beings, a and b are treated as having a well-being of 0 in W_1 . So a 's peak well-being is +5, and b 's peak well-being is 0. In W_1 , there are no individuals, and thus there is no one whose well-being is below their peak. Thus no harm is brought about by the W_1 -option. In W_2 a 's well-being is 0 units below her peak, while b 's well-being is 5 units below her peak. So the harm done by the W_2 -option is: $0 + 5 = 5$. Since the W_1 -option does 0 units of harm and the W_2 -option does 5, HMV prescribes the W_1 -option.

Many people have asymmetric intuitions regarding the moral significance of creating future people.¹⁴ On the one hand, it seems like there's no moral pressure to create more people who would have worthwhile lives. On the other hand, it seems like there is moral pressure to not create people who would have lives not worth living. HMV's method of assessing options captures this asymmetry.¹⁵

Consider a choice between two outcomes, W_1 and W_2 . Subjects with a positive well-being who only exist at W_1 won't make the W_1 -option any more attractive. They'll have their peak well-being at W_1 , and so they won't affect the harm done by the W_1 -option. So the fact that bringing W_1 about will create happy people doesn't give us a reason to bring it about. But subjects with a negative well-being who only exist at W_1 will make the W_1 -option less attractive. Their well-being at W_1 will be below their peak (0), and thus they will increase the harm done by the W_1 -option. So the fact that bringing W_1 about will create unhappy people *does* give us a reason to not bring it about.

4 The Non-Identity Problem and Saturating Counterpart Relations

Consider Parfit's Case of the 14-Year Old Girl:

¹³If one holds the view that all moral agents are moral patients, then this case is, strictly speaking, impossible. I.e., since there is no individual present in all of the outcomes, there couldn't be an agent who was facing these choices. (Recall that these outcomes include all of the agents who exist, at all times.) However, nothing of substance hangs on this, so I'll occasionally engage in the simplifying fiction of ignoring the presence of the agent in question.

¹⁴For example, see Narveson (1967), Wolf (1997), Parsons (2002) and Roberts (2003a).

¹⁵That said, some have argued that this asymmetry is actually counterintuitive. I discuss these arguments in section 8.

A young girl decides to have a child at the age of 14. Because she cannot care for it effectively, the child ends up having a hard life, though a life still worth living. If she had decided not to have a child at the age of 14, but had waited until she was 21, she would have been able to care for the child effectively, and it would have had a much better life.

Even if the girl herself would have been no better off having the child later, it seems clear that what the girl did was wrong. But why? As Parfit notes, our instinctive explanation is a person-affecting one:

“The objection to this girl’s decision is that it will probably be worse for her child. If she waited, she would probably give him a better start in life.” (Parfit (1984), p.359)

With this in mind, we might represent the Case of the 14-Year Old Girl in the following way:

Has Child Now		Has Child Later	
Mother +10	Child +5	Mother +10	Child +10

If we apply a person-affecting view like HMV to this case, we’ll get the result that the latter option is obligatory. In the latter case, both the mother and the child have their peak well-being, while in the former case the child has a well-being 5 units below its peak. Since having the child now will do 5 units of harm, and having the child later will do none, the girl should have the child later.

But, Parfit argues, this way of thinking about the case is mistaken. The child that would be born to the girl at the age of 14 would not be the same as the child that would be born to her at the age of 21, so we can not claim that she has harmed the very *same* child by bringing it into existence now. So our instinctive person-affecting explanation can’t be right. The right way to think about the case, Parfit maintains, is this:

Has Child Now		Has Child Later	
Mother +10	Child ₁ +5	Mother +10	Child ₂ +10

where neither child is a counterpart of the other. And if we apply a person-affecting view like HMV to this case, we’ll get the result that both options are permissible. In both cases, the mother and the child have their peak well-being. So neither option does any harm, and the girl is free to do as she likes.

So HMV’s prescriptions will depend on what counterpart relation we employ. Which counterpart relation should we employ when making moral judgments of this sort?

4.1 Counterpart Relations and Moral Judgments

On counterpart theory, the counterpart relation is picked out by context. But different proponents of counterpart theory might adopt different accounts as to which counterpart

relations are picked out by which contexts. Consider one account—that suggested by the writings of David Lewis.¹⁶ On this account, the counterpart relation delivered by a context is roughly the one that matches our intuitive judgments regarding how to identify subjects in different possibilities in that context. Call this the *Lewisian counterpart relation*.

On Lewis' account, counterpart relations are notoriously context sensitive. If we employ Lewisian counterpart relations to ground moral claims, we risk making our moral claims context sensitive in the same way.

For example, consider Parfit's Case of the 14-year Old Girl. In some contexts—when someone is arguing that if she has her child later then it will be better off, say—the Lewisian counterpart relation may be one that identifies the child she would have now with the child she would have when she's 21.¹⁷ In other contexts—when someone is appealing to the essentiality of origins in order to argue that the child she would have now and the child she would have when she's 21 are not the same, say—the Lewisian counterpart relation may be one that doesn't identify the children in the two cases.¹⁸ These results aren't in conflict. It's just that different contexts will pick out different Lewisian counterpart relations, even when we're considering what is (intuitively) the same case.

Here are two ways one might proceed in light of this. First, one might conclude that, given a person-affecting view, moral claims themselves must be context dependent. And thus, given a person-affecting view, it will turn out that moral claims are not objective in some of the ways we originally thought.¹⁹ This option holds on to the thought that we should employ the Lewisian counterpart relation when making moral judgments, but gives up on the thought that moral claims are objective in all of the ways we thought they were.

Second, one might conclude that while Lewisian counterpart relations are highly context sensitive, the way in which we pair individuals when assessing moral claims is not. This option holds on to the thought that moral claims are objective, but gives up on the thought that we should employ the Lewisian counterpart relation when assessing moral claims.

The first option has some uncomfortable consequences. I take it, for example, that we would like there to be a definite (context-independent) answer to the question of whether it's permissible in the Case of the 14-Year Old Girl for the girl to have the child now. But if we adopt the first option, this will not be the case. Thus I suggest we adopt the second option.

¹⁶For example, see Lewis (1986).

¹⁷On these kinds of questions, Lewis writes: "You could do worse than plunge for the first answer to come into your head, and defend that strenuously. If you did, your answer would be right. For your answer itself would create a context, and the context would select a way of representing, and the way of representing would be such as to make your answer true. ... That is how it is in general with dependence on complex features of context. There is a rule of accommodation: what you say makes itself true, if at all possible, by creating a context that selects the relevant features so as to make it true." Lewis (1986), p.251.

¹⁸"In parallel fashion, I suggest that those philosophers who preach that origins are essential are absolutely right—in the context of their own preaching. They make themselves right: their preaching constitutes a context in which *de re* modality is governed by a way of representing (as I think, by a counterpart relation) that requires match of origins." Lewis (1986), p.252.

¹⁹I use the term "objective" here broadly (if loosely) to cover the rejection of any number of ways in which moral claims might be defective, relative, insubstantial, etc.

There are two different ways to flesh out the second option. One way is to hold one's person-affecting view fixed and to change one's account of which counterpart relations are picked out in moral contexts. On this approach, one will replace the Lewisian counterpart relation with a more stable counterpart relation when evaluating moral claims.

The other way to flesh out the second option is to stick with the Lewisian counterpart relation in moral contexts and modify one's person-affecting view. On this approach, the person-affecting view will not employ counterpart relations. Instead, it will employ some other relations—call them counterpart* relations. These counterpart* relations will presumably line up with counterpart relations in most ordinary contexts, but the two will sometimes come apart. So while it is counterpart relations that determine the truth values of *de re* modal claims, it is counterpart* relations that we employ when applying our person-affecting view. (One might complain that the resulting view is not a person-affecting view, just a person-affecting-ish view. This is not an unreasonable complaint. But if it can solve Parfit's problems then it's an interesting view, regardless of what we decide to call it.)

One can understand the proposals made in this paper either way. Those who think we should employ Lewisian counterpart relations in all contexts have a reason to prefer the second approach. Those who want a view that is full-bloodedly *person-affecting* have a reason to prefer the first approach. But since nothing about these proposals requires me to make a choice, I'll leave it open. To avoid cumbersome repetition, I will continue to talk in terms of counterparts instead of 'counterparts/counterparts*' in what follows.

Let's call a proposal regarding which counterpart relations we should employ when assessing moral claims a *moral-counterpart proposal*. I suggest that we evaluate moral-counterpart proposals according to three desiderata. (i) Stability: we should favor moral-counterpart proposals that employ counterpart relations that are context-insensitive. (ii) Plausible Identifications: we should favor moral-counterpart proposals that match our intuitive judgments regarding how to identify subjects. (iii) Plausible Prescriptions: we should favor moral-counterpart proposals that yield plausible prescriptions when plugged into the correct moral theory.²⁰

Of course, assessing the third desiderata is tricky. We're trying to determine what the right moral-counterpart proposal is by looking at whether it yields plausible prescriptions when plugged into the right moral theory. But we're also trying to determine what the right moral theory is by looking at whether it yields plausible prescriptions when paired with the right moral-counterpart proposal. This puts us in a delicate situation. We're trying to figure out what the right moral theory is and what the right moral-counterpart proposal is at the same time. But our evaluation of each depends on what decisions we make with respect to the other.

As a result, it's hard to evaluate the plausibility of person-affecting views and moral-counterpart proposals in isolation. In order to get a grip on the plausibility of these accounts, we need to assess them in pairs. This, I suggest, is the right way to evaluate the

²⁰If we understand these as desiderata for counterpart*-fixing proposals, it should be clear why we want the first and third desiderata. Why do we want the second desiderata? Because we are, in part, trying to capture person-affecting intuitions. The more a counterpart*-fixing proposal diverges from our intuitive judgments regarding how to identify individuals, the less faithful it is to our person-affecting intuitions.

two proposals being offered in this paper. Instead of trying to evaluate the plausibility of HMV and moral-counterpart proposals separately, we should assess them as a pair.

4.2 Saturating Counterpart Relations

To get specific prescriptions, we need to pair HMV with a moral-counterpart proposal. In what follows, I will tentatively propose a moral-counterpart proposal for us to use.

Let us say that two individuals are *indiscernable-up-to-t* iff they are alike with respect to all of the intrinsic and extrinsic properties that supervene on the qualitative state of the world up to t .²¹ Let us call the worlds that could result from the options available to an agent the agent’s *available worlds*.

Now consider an agent in a decision situation at time t . And consider a counterpart relation which, for each ordered pair of available worlds (W_i, W_j) ($i \neq j$), maps individuals in W_i to counterparts in W_j in a way that satisfies the following four conditions:²²

1. **One-to-One Function:** No individual in W_i is mapped to more than one individual in W_j , and no individuals in W_i are mapped to the same individual in W_j .²³
2. **Before-t Match:** Each individual a who exists before t in W_i is mapped to an individual b who exists before t in W_j that is *indiscernable-up-to-t* with a .
3. **Saturation:** As many individuals in W_i are mapped to individuals in W_j as possible.
4. **Minimization:** There is no mapping which satisfies the first three conditions and which results in the W_i -option having a lower harm.

²¹In relativistic worlds we can instead consider what it is for two individuals to be *indiscernable-up-to-r*, where r is the spatiotemporal region the agent occupies at the moment of decision. We can say that two individuals are *indiscernable-up-to-r* iff they are alike with respect to all of the properties and relations that supervene on the qualitative state of the world in the backwards light cone of r . (I’m assuming here that there aren’t closed timelike curves; some other strategy needs to be employed if there are.)

²²My use of the term “maps” should be understood to imply only that there is a multivalued function (or “multimap”) from individuals in W_i to individuals in W_j , not that there is a function (or “map”) from the former to the latter. The first condition below will, in fact, require there to be such a function. But we want all of the substantive constraints on the counterpart relation to appear in the list of conditions, not to be smuggled in by our set-up.

²³If we take counterpart relations to be similarity relations, then this condition is a bit too strong. Problems arise in cases in which there are multiple indiscernable individuals at a world—individuals who share *all* of their intrinsic and extrinsic qualitative properties. Because these individuals are indiscernable, a qualitative counterpart relation can’t assign them different counterparts, or take them to be counterparts of different individuals. So in these cases both parts of condition 1 can fail.

There are a couple of different ways to handle such cases. One approach is to shift the qualitative requirement from the counterpart relations themselves to what counterpart relations a context can pick out. Then we could allow counterpart relations to be more fine-grained (and so allow them to be one-to-one functions even in cases with multiple indiscernable individuals), but require contexts to deliver multiple counterpart relations—all of the counterpart relations that are ‘precisifications’ of the coarse counterpart relation the original theory employed. Then, when assessing person-affecting views like HMV, one could employ any or all of these fine-grained counterpart relations, since they’ll all deliver the same results.

Since the most distinctive feature of these counterpart relations is provided by the third condition, I'll call a counterpart relation which satisfies these four conditions a *saturating counterpart relation*.

These four conditions won't pick out a unique counterpart relation. For example, at worlds in which there are multiple individuals that are indiscernable-up-to- t , there will often be wiggle room with respect to which one serves as the counterpart of some qualitatively similar other-worldly individual.²⁴ For a more mundane example, if there are multiple individuals at a world who come into existence after t and have the same level of well-being, then counterpart relations which permute them will satisfy these conditions equally well. But this wiggle room needn't bother us. All of the counterpart relations which satisfy these four conditions will yield the same prescriptions when coupled with HMV. So it doesn't matter which one we use. (This is the main reason for including the fourth condition—it ensures that any wiggle room that remains won't bear on HMV's prescriptions.)

I propose to pair HMV with the following moral-counterpart proposal: when making moral judgments, we should employ counterpart relations which satisfy conditions 1-4; i.e., saturating counterpart relations. In the previous section I offered three desiderata for assessing a moral-counterpart proposal for a person-affecting view: stability, plausible identifications, and plausible prescriptions. I think the moral-counterpart proposal given by conditions 1-4 does a good job of satisfying these desiderata when paired with HMV. It satisfies stability, it does relatively well with respect to plausible identifications, and (as I'll argue) it does well with respect to plausible prescriptions. Now, one could do better with respect to plausible identifications by adding some additional 'matching' conditions which are assessed before the fourth condition. And some of these modified proposals will do just as well, if not better, with respect to plausible prescriptions. So I won't claim that conditions 1-4 yield the *optimal* moral-counterpart proposal. But I will claim that in most cases conditions 1-4 will yield the same prescriptions as the optimal moral-counterpart proposal. So I think conditions 1-4 do well enough to allow us to fairly assess the merits and demerits of this approach to person-affecting views.

Let's call the combination of this moral-counterpart proposal and HMV the *Saturating Harm Minimizing View* (SHMV).

A warning: we can't just map an individual at an available world W to all of the other available worlds in a way that satisfies these conditions, and then take all of these individuals to be counterparts of one another. The reason is that the counterpart relation is generally not symmetric or transitive, and *a fortiori* is generally not an equivalence relation. So when we're evaluating whether a counterpart relation satisfies these conditions, we need to assess *each* ordered pair of available worlds.

Similar remarks apply to the manner in which we assess the harm of an option. Here

²⁴All of an agent's outcomes will be identical up to t . So if there are multiple individuals that are indiscernable-up-to- t at one available world, there will be the same number of individuals who are indiscernable-up-to- t at every other available world. Thus there may be looseness regarding which of these indiscernable-up-to- t individuals are mapped to which other indiscernable-up-to- t individuals. (Even this looseness will sometimes be removed by the fourth condition, if the individuals end up having different levels of well-being due to their experiences after t , and this ends up impacting the harm assigned to the world.)

is how to assess the harm of some W -option with respect to a given counterpart relation. First, for each individual in W , determine who they're mapped to in every available world. Second, determine the peak well-being of each individual in W by finding which of these counterparts has the highest well-being (where they're treated as having a counterpart with a well-being of 0 at worlds in which they don't have counterparts). Third, consider how much the well-being of each individual in W falls below their peak, and sum these values. The resulting quantity is the harm brought about by the W -option. And to assess the harm of our other options, we must go through the same procedure.

Example: The Somewhat-Happy Addition. Consider an agent who has a choice between the following two outcomes:

W_1	W_2	
a	b	c
+10	+10	+5

Let's begin by determining what the saturating counterpart relations are, and then determine how much harm is done by each option.

First let's consider who the individuals in W_1 will be mapped to. Suppose that none of these subjects exist at the time of the choice, so the before- t match condition doesn't come into play. The saturation condition requires a to be mapped to either b or c . In either case a 's peak well-being will be +10, and so the W_1 -option will do no harm; thus either mapping will satisfy the minimization condition. So a can be mapped to either b or c .

Next, let's consider who the individuals in W_2 will be mapped to. The saturation condition requires either b or c to be mapped to a . If b is mapped to a , then b 's peak well-being will be +10, c 's peak-well being will be +5, and the W_2 -option will do no harm. If c is mapped to a , then both b and c 's peak well-being will be +10, and the W_2 -option will do 5 units of harm. So the minimization condition requires b to be mapped to a .

Given these mappings, neither option does any harm. So SHMV takes both options to be permissible.

4.3 The Non-Identity Problem

Let us return to the Case of the 14-Year Old Girl. Parfit suggests that we think of the case like this:

Has Child Now		Has Child Later	
Mother	Child ₁	Mother	Child ₂
+10	+5	+10	+10

where neither child is a counterpart of the other. As we saw in section 4, if Parfit is right about how we should identify subjects when making person-affecting judgments, then HMV yields the counterintuitive result that there's nothing wrong with the girl having the child now.

But if Parfit is right, we're left with a puzzling question. Why do we have the instinctive reaction to this case that Parfit describes? Why does it seem to us that "the objection to this girl's decision is that it will probably be worse for her child... if she waited, she

would probably give him a better start in life”?²⁵ This seems to be a paradigmatic case of a person-affecting judgment.²⁶ But its hard to reconcile this judgment with the claim that we should use the counterpart relation Parfit suggests when making person-affecting judgments.

Let’s consider a different approach. Suppose, as I’ve suggested, that we’re inclined to pair up as many subjects as possible when making moral judgments. I.e., suppose that we make moral judgments using a *saturating* counterpart relation. The mothers will be mapped to one another because they are indiscernable-up-to-*t*, and the saturation condition will then require the two children to be mapped to one another. Thus we’ll represent the Case of the 14-Year Old Girl in the following way:

Has Child Now		Has Child Later	
Mother	Child	Mother	Child
+10	+5	+10	+10

And as we saw in section 4, giving this pairing of subjects, HMV will yield the desired result that having the child now is morally impermissible.²⁷

This justifies our initial reaction—that the first option is worse because it is worse *for* the child.²⁸ It’s true that the two children are not counterparts according to the counterpart relation Parfit suggests. But that is not the counterpart relation that we should use when making moral judgments. The right counterpart relation to use is a saturating one. And when we use a saturating counterpart relation, HMV yields the moral judgment that we’re initially inclined to give.

5 The Repugnant Conclusion

Consider Parfit’s Repugnant Conclusion: “For any possible population of at least ten billion people, all with a very high quality of life, there must be some much larger imaginable population whose existence, if other things are equal, would be better, even though its members have lives that are barely worth living.”²⁹ To simplify a bit, suppose we have a choice between two options, which lead to the following outcomes:

W_1	W_2
$a_1 - a_{10}$	$b_1 - b_n$
+100	+1

²⁵Parfit (1984), p.359.

²⁶One might suggest that we understand the assertion that “this girl’s decision... will probably be worse for her child” as a *de dicto*, not a *de re* claim (see Hare (2007)). If so, then this is not a person-affecting judgment, and talk of counterpart relations is besides the point. I think the second half of this assertion—“if she waited, she would probably give *him* a better start in life”—suggests a *de re* reading. But in any case, not much hangs on this. SHMV delivers the correct prescription regardless of what story we end up deciding on.

²⁷A similar response to the Non-Identity Problem is suggested by Wrigley (2006), who employs counterpart theory to assess the moral status of genetic selection.

²⁸If we employ counterpart* relations we may have to hedge this claim a bit, since there are contexts in which the counterpart and counterpart* relations can come apart.

²⁹Parfit (1984), p.388.

Further suppose, as Parfit suggests, that there are entirely different populations in W_1 and W_2 ; we are like deities choosing to create one of two very different universes. That is, suppose that none of the individuals in either world is a counterpart of an individual in the other (putting aside, for the moment, the moral counterpart proposal of section 4.2). If there is some n large enough to make the W_2 -option obligatory, we're led to the Repugnant Conclusion.³⁰ Does HMV yield this result?

No. None of these individuals have any counterparts in the other world. So all of these individuals are at their peak well-being. Thus neither option does any harm. And HMV will take both options to be permissible, regardless of how large n is. So HMV avoids the Repugnant Conclusion.

That said, this way of avoiding the Repugnant Conclusion isn't very satisfying. Let's distinguish between the Strong Repugnant Conclusion—that the W_2 -option is obligatory, and the Weak Repugnant Conclusion—that the W_2 -option is permissible. Parfit identifies the Repugnant Conclusion with the strong version. And HMV avoids this conclusion by taking the W_1 and W_2 -options to be on a par. But most people feel that not only is the W_2 -option not obligatory, the W_2 -option is not even permissible. And since HMV takes both the W_1 and W_2 -options to be permissible, HMV does not capture this intuition.

Let me suggest an explanation for why the W_2 -option strikes us as strictly worse than the W_1 -option.³¹ Our moral judgments tend to be comparative in nature. We try to assess the importance of the well-being of different subjects in comparative terms as much as possible. And although we've been told in the above case that none of the subjects in the two outcomes correspond to the same individual, we're still inclined to pair up as many of them as possible for the purposes of comparison.

If this explanation is correct, then there is a natural way to capture the intuition that the W_2 -option is impermissible. We can employ a counterpart relation which pairs as many subjects in different outcomes as it can; i.e., a *saturating* counterpart relation. Then, in the above case, we can map the ten subjects in W_1 to ten of the subjects in W_2 and vice versa, and think about the case like this:

W_1	W_2	
$a_1 - a_{10}$ +100	$a_1 - a_{10}$ +1	$b_{11} - b_n$ +1

Given this saturating counterpart relation, HMV yields the desired result that only the W_1 -option is permissible. In W_1 , $a_1 - a_{10}$ have their peak well-being. So the W_1 -option does no harm. In W_2 , $b_{11} - b_n$ have their peak well-being, but $a_1 - a_{10}$ have a well-being 99 units below their peak. So the W_2 -option brings about $10 \times 99 = 990$ units of harm. Since the W_1 -option brings about 0 units of harm while the W_2 -option brings about 990, the W_1 -option is obligatory.

³⁰Strictly speaking, Parfit is talking about assessments of which worlds are better than one another, not assessments of what one ought to do. But I take the interesting question to be the one concerning obligation; questions regarding which world is better are only interesting insofar as they relate to what we ought to do. So this is how I'll understand the problems Parfit raises. (See also the discussion in section 7.)

³¹Of course, nothing much hangs on this explanation. SHMV yields the right result regardless of whether this explanation of our intuitions is correct.

So SHMV avoids both the strong and weak versions of the Repugnant Conclusion.

6 The Absurd Conclusion

Consider Parfit's Absurd Conclusion: there can be a moral difference between worlds whose populations have the same distributions of well-being, but where the subjects live concurrently instead of consecutively. So suppose we have a choice between two options. One option leads to a "concurrent world", a world in which there are $n > 1$ individuals, each with a well-being of m , who come into existence at the same time and die off at the same time. The other option leads to a "consecutive world", a world in which there are also n individuals with a well-being of m , but where each comes into existence alone and dies off before the next individual is created. If there is some n and m which makes only one of these options permissible, we're led to the Absurd Conclusion.

Will SHMV lead to this conclusion? No. To see why, let's work out what the saturating counterpart relations between these two outcomes will be.

First, note that both of these populations will consist of future people, so the before- t match condition will not apply. (When an agent faces a choice at t , all of her potential outcomes will be identical up to t . So if a concurrently existing population has already existed before t , there will be concurrently existing individuals in every outcome, including the consecutive world. Likewise, if any lonely individuals have already existed before t , there will be lonely individuals in every outcome, including the concurrent world.) The saturation condition requires each of the subjects in each world to be mapped to a subject in the other. And the minimization condition requires agents with the same well-being to be mapped to each other, since these are the mappings that will minimize the harm of each option.

So the saturating counterpart relation will map all the individuals in each world to counterparts in the other who have the same well-being. Every individual in both worlds will have their peak well-being, and neither option will do any harm. Thus SHMV will take both options to be permissible.

7 The Mere-Addition Paradox and the Independence of Irrelevant Alternatives

Suppose that an agent faced with a choice between outcomes W_1 and W_2 would prefer the W_1 -option to the W_2 -option. Then it seems she should continue to prefer the W_1 -option to the W_2 -option when faced with a choice between W_1 , W_2 and some third outcome, W_3 . After all, it's hard to see why the inclusion of this third outcome should bear on the relative merits of W_1 versus W_2 . More generally, it seems that an agent's preferences regarding a W_1 -option versus a W_2 -option should be independent of what other options are available. This requirement is a version of the "Independence of Irrelevant Alternatives" (IIA), one of the canonical decision-theoretic constraints on the preferences of rational agents.

(If we accept IIA, and assume that rational agents can have preferences which line up with the “all-things-considered-better-than” relation, then IIA will also constrain this “better than” relation. I take it that rational agents can have preferences which line up with the “all-things-considered-better-than” relation. Thus, although I speak in terms of preferences in what follows, what I say applies *mutatis mutandis* to the “all-things-considered-better-than” relation.)

Strictly speaking, IIA doesn’t say anything about normative theories like SHMV. Normative theories like SHMV are accounts about what options are morally permissible, not about what preferences one should have. But there’s a natural way to link normative theories to preference constraints like IIA. Let’s say that a normative theory *meshes* with a preference constraint *C* iff an agent who always prefers the options prescribed by the theory can satisfy *C*. Then we can bring IIA to bear on a normative theory by asking whether the theory meshes with IIA.

We can formulate the requirement that a theory mesh with IIA in deontic terms. Call a decision situation in which both a W_1 -option and a W_2 -option are available a W_1W_2 -situation. Then we can formulate the requirement as follows:

Deontic IIA (IIA_d):

- (i) If there exists a W_1W_2 -situation in which both the W_1 and W_2 -options are permissible, then in all W_1W_2 -situations the W_1 -option is permissible iff the W_2 -option is permissible.
- (ii) If there exists a W_1W_2 -situation in which the W_1 -option is permissible and the W_2 -option is impermissible, then in all W_1W_2 -situations the W_2 -option is impermissible.

A normative theory meshes with IIA iff it satisfies IIA_d.³² (The proof is provided in the appendix.)

IIA_d seems like a plausible constraint. However, SHMV appears to violate IIA_d. To see this, consider two decisions. First consider the choice between the following two outcomes:

W_1	W_2
	<i>a</i> +5

The W_1 -option doesn’t do any harm to anyone, since there’s no one in W_1 . The W_2 -option doesn’t do any harm either, since *a* has her peak well-being in W_2 . So both options are permissible.

Now suppose we add a third outcome, W_3 :

W_1	W_2	W_3
	<i>a</i> +5	<i>a</i> +10

³²Interesting questions arise regarding how to understand conditional deontic claims if we reject IIA_d. (Thanks to Ted Sider here.) Although these are interesting issues, I won’t attempt to address them here.

Again, the W_1 -option won't do any harm. And the W_3 -option won't do any harm either, since a has her peak well-being in W_3 . But the W_2 -option will now do 5 units of harm, since a 's well-being in W_2 is 5 units below her peak. Thus only the W_1 and W_3 -options are permissible.

But this appears to violate IIA_d . Both of these cases are W_1W_2 -situations. And since the W_1 and W_2 -options are both permissible in the first case, IIA_d requires a W_2 -option to be permissible whenever a W_1 -option is. But in the second case, the W_1 -option is permissible and the W_2 -option is not.

Should we take this to be a reason to reject SHMV? Here are two reasons to think not.

First, as Roberts (2003b) points out, it's not clear that person-affecting views like SHMV actually do fail to satisfy principles like IIA_d . If we think of W_1 and W_2 in an appropriately detailed way, then the outcomes in the two cases won't be the same. In the first case, the outcome we called " W_1 " will include facts about an agent who faced a choice between *two* outcomes, while in the second case, the outcome we called " W_1 " will include facts about an agent who faced a choice between *three* outcomes. And if these are different outcomes, then these two cases will involve different decision situations: in the one case we have a W_1W_2 -situation, in the other a $W_1^*W_2^*$ -situation. Since IIA_d only places constraints on prescriptions for situations of the same kind, SHMV's prescriptions in these two cases won't have any bearing on each other. So given this detailed picture of outcomes, SHMV will satisfy IIA_d .

Of course, if we think of outcomes as being detailed in this way, then it will be impossible to have preferences that violate IIA_d , since the same outcome will never appear in different decision situations. This makes principles like IIA and IIA_d vacuous. One might take this to be a reason to think of the outcomes we're considering in a more coarse-grained way.

For the sake of argument, I will grant the kind of coarse conception of outcomes required to make principles like IIA_d non-trivial in what follows. Likewise, I will grant that person-affecting views like SHMV will not satisfy IIA_d .

Let's turn to the second reason for not rejecting SHMV. Although the violation of IIA_d first seems like a demerit of the account, there are reasons to think that it is in fact a strength. As we will see, it is this violation of IIA_d that allows SHMV to satisfy our intuitive judgments in the cases comprising Parfit's Mere-Addition Paradox. Indeed, given the assumption that there must always be a permissible option available, we'll see that this violation is inescapable: *any* solution which captures all of these intuitions *must* violate IIA_d .

Let's examine each of these points more carefully.

7.1 The Mere-Addition Paradox

Consider the three cases that lead to the Mere-Addition Paradox.

First, consider a choice between the following two outcomes:

W_1	W_2	
$a1 - a10$	$b1 - b10$	$b11 - b20$
+10	+10	+5

It seems like W_2 must be at least as good as W_1 . After all, the same number of equally well-off subjects exist, and then there are some additional well-off subjects hanging around as well. So intuitively, either both the W_1 and W_2 -options are permissible, or only the W_2 -option is permissible.

Second, consider a choice between the following two outcomes:

W_2		W_3	
$b1 - b10$	$b11 - b20$	$c1 - c10$	$c11 - c20$
+10	+5	+9	+9

It seems like W_3 must be better than W_2 . There are the same number of people in both, and the people are significantly happier on average in W_3 than they are in W_2 . So intuitively, only the W_3 -option is permissible.

Third, consider a choice between the following two outcomes:

W_1	W_3	
$a1 - a10$	$c1 - c10$	$c11 - c20$
+10	+9	+9

It seems like W_1 is at least as good as W_3 . Indeed, if we increase the disparities in the number of agents and their well-being's in the different outcomes in these cases, this case turns into the Repugnant Conclusion case discussed in section 5, where it's clear that W_1 is better than the alternative. So intuitively, either both the W_1 and W_3 -options are permissible, or only the W_1 -option is permissible.

As Parfit (1984) noted, these three judgments appear to be in tension. One way to characterize this tension is in terms of preferences. Let " \geq " stand for the "preferred at least as much as" relation, and " $>$ " stand for the "preferred more than" relation. Then these judgments suggest preferences according to which a W_1 -option \leq a W_2 -option $<$ a W_3 -option \leq a W_1 -option. But this ranking is incoherent, since it requires the W_1 -option to be preferable to itself.

However, it's easy to be distracted by tangential matters when we characterize the issue in terms of preferences.³³ We can avoid these distractions by characterizing the tension as a straightforward contradiction in deontic terms. Namely, given IIA_d and the assumption that some option must be permissible, these three prescriptions lead to a contradiction.

The full proof is given in the appendix. But let's see how to get the contradiction given the most natural judgments in these cases: that both options are permissible in the first case, that only the W_3 -option is permissible in the second case, and that only the W_1 -option is permissible in the third case.

Consider a choice between all three of the above outcomes:

W_1	W_2		W_3	
$a1 - a10$	$b1 - b10$	$b11 - b20$	$c1 - c10$	$c11 - c20$
+10	+10	+5	+9	+9

³³Boonin-Vail (1996) and Arrhenius (2004) are among those who suggest that these issues are better evaluated by characterizing the paradox in deontic terms instead of a 'better-than' or preference ranking.

Given the natural judgment in the choice between W_1 and W_2 —that both options are permissible—I A_d entails that the W_1 -option is permissible *iff* the W_2 -option is permissible. Given the natural judgment in the choice between W_2 and W_3 —that only the W_3 -option is permissible—I A_d entails that the W_2 -option is impermissible. Since we’ve seen that the W_1 -option is permissible *iff* the W_2 -option is permissible, it follows that the W_1 -option is impermissible as well. Finally, given the natural judgment in the choice between W_1 and W_3 —that only the W_1 -option is permissible—I A_d entails that the W_3 -option is impermissible. So all three options are impermissible. But some option must be permissible. Contradiction.³⁴

7.2 SHMV and the Mere-Addition Paradox

How does SHMV deal with this Paradox? To find out, let’s look at what SHMV says about each of the cases that comprise the Mere-Addition Paradox.

Consider the first case:

W_1	W_2	
$a1 - a10$	$b1 - b10$	$b11 - b20$
+10	+10	+5

This case is identical to *The Somewhat-Happy Addition* case discussed in section 4.2, except for the fact that there are ten times as many subjects. Multiplying the number of subjects in a uniform way like this won’t change SHMV’s prescriptions, however. So SHMV will yield the same verdict as before: both options are permissible.

Now consider the second case:

W_2		W_3	
$b1 - b10$	$b11 - b20$	$c1 - c10$	$c11 - c20$
+10	+5	+9	+9

Since the number of individuals in both outcomes is the same, the saturation condition requires us to map each individual in one outcome to an individual in the other. And since all of the individuals in W_3 have the same well-being, it won’t matter what mapping we choose. So suppose we pair the subjects in each outcome in numerical order (i.e., $b1$ with $c1$, $b2$ with $c2$, etc.). Then the first ten subjects will have a peak well-being of +10, and the second ten subjects will have a peak well-being of +9. It follows that the W_2 -option will do 40 units of harm, while the W_3 -option will do 10 units of harm. Thus the W_3 -option is obligatory.

Finally, consider the third case:

³⁴A number of results demonstrating the incompatibility of several normative theses that yield these three judgments have been given in the literature; see Ng (1989), Blackorby and Donaldson (1991), and Arrhenius (2000). The result stated here, and proved in the appendix, is both weaker and stronger than these results. It is stronger in that it makes no assumptions about the normative theses that justify our intuitive judgments in these three cases, and thus applies regardless of how one tries to justify these verdicts. It is weaker in that it doesn’t directly yield conclusions regarding which kinds of normative theses are mutually inconsistent. (Though one can use this result to generate such conclusions by finding sets of principles that yield the three verdicts in question.)

W_1	W_3	
$a1 - a10$	$c1 - c10$	$c11 - c20$
+10	+9	+9

The saturation condition requires us to map ten individuals in W_1 to individuals in W_3 , and vice versa. And since all of the individuals in W_3 have the same well-being, it won't matter what mapping we choose. So suppose we pair the first ten subjects in each outcome. Then the peak well-being for the first ten subjects will be +10, and the peak well-being for the other ten subjects will be +9. It follows that the W_1 -option will do no harm, while the W_3 -option will do 10 units of harm. Thus the W_1 -option is obligatory.

So SHMV yields the same verdicts as our intuitive judgments do in the first three cases. But how, then, does it avoid the contradiction that these judgments lead to? To see, let's consider how SHMV treats the case in which all three outcomes are available:

W_1	W_2		W_3	
$a1 - a10$	$b1 - b10$	$b11 - b20$	$c1 - c10$	$c11 - c20$
+10	+10	+5	+9	+9

One can show that pairing the first ten subjects in each outcome, and the next ten subjects in W_2 and W_3 , is a saturating counterpart relation. Given this pairing, the peak well-being for the first ten subjects will be +10, and the peak well-being for the other ten subjects will be +9. It follows that the W_1 -option does no harm, the W_2 -option does 40 units of harm, and the W_3 -option does 10 units of harm. Thus the W_1 -option is obligatory.

So SHMV avoids the contradiction. And it does so by violating IIA_d : given SHMV's prescriptions in the first three cases, IIA_d requires SHMV to maintain that the W_1 -option is impermissible in the combined case. But SHMV maintains that the W_1 -option is permissible.

Although this violation of IIA_d initially looked like a weakness of the SHMV, we can now see that it is a strength. Given that some option is always permissible, the only way to capture our intuitive judgments in the three cases is to reject IIA_d . So in order to offer an intuitively satisfying response to the Mere-Addition Paradox, IIA_d must be rejected.³⁵

This puts us in a position to see why the various kinds of "impossibility theorems" that have been offered in the literature—results showing that no theory can satisfy all of

³⁵Another principle along these lines that person-affecting views conflict with is the *Pareto Plus Principle* (PPP): if a W_1 -option is permissible, a W_2 -option is available, and W_2 is the same as W_1 except that it contains an additional happy person, then the W_2 -option must be permissible. I don't think this conflict raises any additional interesting issues, however. Rather, I think that the conflict between person-affecting views and PPP is just the conflict between person-affecting views and IIA_d in disguise.

To see why, consider the *Restricted Pareto Plus Principle* (RPPP), which applies solely to cases in which there are only two options available. I suggest that RPPP captures the distinctive intuition behind PPP. And person-affecting views like SHMV won't conflict with RPPP. But we can derive PPP from RPPP if we assume IIA_d . And person-affecting views like SHMV will conflict with PPP. So it isn't until we add IIA to RPPP that we get a conflict with SHMV. This suggests that the conflict between person-affecting views like SHMV and PPP stems from the implicit IIA -like assumptions built into the formulation of PPP, not from anything distinctive regarding PPP *per se*. (See Roberts (2003b) for another argument for why proponents of person-affecting views should reject PPP.)

some set of desirable features—are not a threat to person-affecting views like SHMV.³⁶ These theorems explicitly assume that outcomes can be ranked according to their value in a situation-independent way. And these theorems implicitly assume that this notion of value is relevant to determining our moral obligations. If this notion of value had nothing to do with what we ought to do, then these results would be of little interest.

But proponents of person-affecting views will take one of these two assumptions to be false. They can grant that there are notions of “value”, such as monetary value, with respect to which the values of outcomes can be ranked in a situation-independent way. But they will deny that these notions of value are morally interesting, since they have little to do with our moral obligations. Likewise, they can grant that there are notions of “value” that track what we ought to do, such as the harm done by the outcome, with respect to which the values of outcomes can be ranked in a given situation. But then they will deny that the value of an outcome can be determined in a situation-independent way, since the harm done by an outcome will depend on what other outcomes are available to the agent in that situation.

This also allows us to see the arguments offered by proponents of intransitivity, such as Temkin (1987), Rachels (1998) and Persson (2004), in a new light. Proponents of intransitivity can be seen as arguing that there is no “all things considered better than” relation which is (i) directly tied to moral obligation, (ii) situation-independent, and (iii) transitive. Proponents of person-affecting views like SHMV will agree. But proponents of intransitivity take the culprit to be (iii). Proponents of person-affecting views will take the culprit to be either (i) or (ii). I.e., either the “all things considered better than” relation is not directly tied to moral obligation (in which case it’s of little interest), or it’s not situation-independent.³⁷

8 Objections

What objections to SHMV might one have? Let me briefly consider five kinds of objections, in ascending order of strength.

First, one might reject counterpart theory. In this paper I’ve simply assumed that counterpart theory is correct. And those who reject counterpart theory might get off the boat right from the start.

That said, it’s worth noting that even those who reject counterpart theory could employ the machinery of SHMV. They could take the algorithm for determining the deontic status of one’s actions that SHMV provides, and strip it of the counterpart theoretic interpretation it’s been given here. Of course, if we pursue this approach, this algorithm is less well-motivated. How heavy this cost is, and whether the results SHMV yields are attractive enough to overcome it, is a question I’ll leave for others to decide.

³⁶For examples of such theorems, see Ng (1989), Blackorby and Donaldson (1991), and Arrhenius (2000).

³⁷What if one thinks that it’s analytic that an “all things considered better than” relation will satisfy (i)-(iii)? Then proponents of person-affecting views will follow proponents of intransitivity in denying that there is such a relation.

A second source of objections stems from the fact that, like the canonical forms of utilitarianism, SHMV is a well-being focused theory. And we have lots of moral intuitions regarding things like rights, justice, desert, equality, and so on, that typical well-being-focused theories don't accommodate.³⁸ There are a couple of ways to respond to these worries within the general framework of this approach. The first is to follow the utilitarian tradition of trying to defuse these kinds of intuitions. The second is to try to incorporate such considerations into the basic notions the account employs. In the case of desert, for example, we might follow Feldman (1995) and incorporate such considerations into our assessment of a subject's well-being. I have little to say about which of these approaches we should employ with respect to which worries. But given the availability of these kinds of responses, I take it that these worries, while interesting and reasonable, do not threaten the viability of person-affecting approaches like SHMV.

A third source of objections comes from disagreements regarding Parfit's desiderata. A number of people have argued that one or more of Parfit's four requirements are misguided, and that once we think about these cases in the right way, we'll see that some of the counterintuitive results Parfit tries to avoid are not really counterintuitive after all. For example, some have argued that we should come to accept the Repugnant Conclusion.³⁹ I don't want to rule out the possibility that these claims are correct.⁴⁰ That said, I think it's clear that there is at least a strong *prima facie* case to be made in favor of these verdicts. So I don't take these to be compelling objections to SHMV.

A fourth source of objections stems from worries regarding the asymmetry regarding the moral significance of creating future people described in section 3.2. For asymmetric theories like SHMV, there is moral pressure to not create individuals with negative well-being, but no corresponding pressure to create individuals with positive well-being. Consider again the Question of Creation case described in section 3.2:

The Question of Creation. Consider an agent who has a choice between two outcomes—creating no one, or creating both a happy person and an unhappy person:

W_1	W_2	
	a	b
	+5	-5

On SHMV, the presence of the +5 individual is not a mark in favor of the W_2 -option, but the presence of the -5 individual is a mark against it. And there is nothing that tells against the W_1 -option. So according to SHMV, it's obligatory to choose the W_1 -option.

In section 3.2 I suggested that this asymmetric treatment of future individuals is intuitively plausible. But some have argued that this asymmetry is actually counterintuitive (see Sikora (1978) and Holtug (2004)). In the Question of Creation example, it is suggested that both options should be permissible. And if we were to make a 's well-being

³⁸Likewise, our moral intuitions may distinguish between things like preventing harms and providing benefits, something that typical well-being-focused theories won't be sensitive too. (Thanks to Elizabeth Harman here.)

³⁹For example, see Sikora (1978), Mackie (1985), Hare (1993), Ryberg (1996), Holtug (2004) Tännsjö (2004), and Huemer (2008).

⁴⁰I say this because I'm sympathetic to these kinds of utilitarian apologetics. Indeed, I think something like utilitarianism may well be correct.

a little higher (+6, say), it is suggested that the W_2 -option should be obligatory. I do not share these intuitions, though my feelings here aren't very strong. But let me note that a different kind of case, which has been taken by critics to provide a decisive objection to the asymmetry, falls short of the task.

Sikora (1978) and Holtug (2004) both discuss the question of whether we should continue to have children and propagate the human race, or whether we should stop reproducing and let the human race fall into extinction. In the former case we will bring about the existence of many more people, most of them happy, but a few with lives not worth living. If we accept the asymmetry, then there's pressure to not create unhappy individuals, but no countervailing pressure to create happy individuals. So it will be better to create no one than to create a bunch of future individuals, a few who would be unhappy. Thus if we accept the asymmetry, the critics argue, we're obligated to stop having children and to let the human race go extinct.

Although our intuitions about this case are much stronger than in the Question of Creation example, I think this is a bad case to appeal to. First, this case drags in a number of misleading or orthogonal intuitions, such as implicit assumptions about the desires of the populace and the consequences of such choices on their well-being, sentiments about things like the "right to procreate", intuitions regarding the intrinsic value of the survival of the species, and so on. (See Wolf (1997) for a discussion of some of these issues.) And these issues are orthogonal to the question of whether or not there's an asymmetry with respect to well-being.

Second, the argument won't generally go through in realistic cases. Consider: why think that your choice to procreate will result in the existence of individuals whose lives are not worth living? The thought might be this: "The effects of your choice to procreate will ripple outward, and change a great many things. And it may result in some individuals being harmed relative to their counterparts in the outcome that results from a different choice." But this is just as true of the choice not to procreate. And there's no reason to think that the decision to procreate will lead to more harm, all things considered, than the decision not to procreate. (Indeed, to get the conclusion that you should never procreate, it needs to be the case that *all* of your options to procreate (at any time, with any partner) will result in more harm than the other options available.)

Third, to the extent that we're concerned with subjective obligations, our assessment of this case will hang on tricky issues regarding probability—issues that we've been avoiding so far. When we choose to have children, we're taking a gamble with respect to how well-off their lives will be. We may be relatively confident that they'll have lives worth living, but we can't be entirely certain of this. In order to argue that people like us are obligated not to have children, given SHMV, the critic needs to claim that the epistemic possibility of our child not having a life worth living is sufficient to make it impermissible to have that child. But whether this is true will depend on how we decide to incorporate uncertainty into our theory. And there are natural ways of doing this—evaluating harm with respect to the *expected* well-being of a subject, for example—which will not yield the claim the critics require.

We can get around these complications by setting up a more straightforward case, such as the following: a deity is able to bring about one of two outcomes, both full of well-off

subjects who will propagate indefinitely. But one outcome contains an additional pair of subjects, one who is extremely well-off (has a well-being as high as you like), and one who is so miserable that her life isn't worth living, though only barely so. This case avoids my complaints. And asymmetric theories like SHMV will maintain that the deity should decline to create the additional pair of subjects. But once we clean up the case like this, I no longer have the intuition that this prescription is incorrect.

A fifth kind of objection stems from cases like the following:⁴¹

Asymmetric Creation. Consider an agent who has a choice between the following three outcomes:

W_1	W_2		W_3	
	a_1	a_2	b_1	b_2
	+5	+10	+6	+9

One can show that a saturating counterpart relation will map a_1 to b_1 and a_2 to b_2 , and vice versa. So the W_1 -option will do no harm, while both the W_2 and W_3 -options will do 1 unit of harm (since a_1 is 1 unit below her peak well-being in W_2 , and b_2 is 1 unit below her peak well-being in W_3). Thus according to SHMV, the W_1 -option is obligatory.

This may seem like a funny prescription for SHMV to make. At first glance, one might think that all three options should be permissible, not just the W_1 -option. What's going on?

Here is my diagnosis. I think IIA_d -style reasoning is illicitly sneaking into our assessment of this case. If we just had the W_1 and W_2 -options to choose between, both options would be permissible. And if we just had the W_1 and W_3 -options to choose between, both options would be permissible. So when presented with a case with all three options available, it's natural to implicitly appeal to IIA_d -reasoning to reach the conclusion that all three options should be permissible.

But as we saw in section 7.2, we should be wary of IIA_d -style reasoning. And such reasoning won't generally lead to the intuitively correct prescriptions in these kinds of cases. Consider the following case:

Dominant Creation. Consider an agent who has a choice between the following three outcomes:

W_1	W_2		W_3	
	a_1	a_2	b_1	b_2
	+5	+10	+10	+20

Regardless of how we map the individuals in W_2 and W_3 to one another, the W_2 -option will do 15 units of harm and the W_3 -option will do no harm. Thus on SHMV both the W_1 and W_3 -options are permissible.

I take it that SHMV delivers the intuitively correct prescription in this case. But this prescription entails the same kind of IIA_d -violation as SHMV's prescription in the previous case. As before, if we just had the W_1 and W_2 -options, or the W_1 and W_3 -options, then both options would be permissible. But we don't want to conclude from this that all three options are permissible in the Dominant Creation case.

⁴¹I owe this case to James Patten.

If my diagnosis of the Asymmetric Creation case is correct, then the Asymmetric Creation case does not yield a problem for SHMV. Instead, it yields a moral: we need to be careful not to slip into IIA_d-style reasoning when evaluating SHMV's prescriptions.

9 Conclusion

I've presented a person-affecting approach to the problems in population ethics that Parfit (1984) raises. The first part of the approach is a particular person-affecting view, the Harm Minimizing View:

The Harm Minimizing View (HMV): An option is morally permissible (for a at t) iff it minimizes harm.

The second part of the approach is a moral-counterpart proposal:

The Moral-Counterpart Proposal: When applying HMV, we should employ saturating counterpart relations.

Together, these two claims comprise the Saturated Harm Minimizing View (SHMV).

SHMV has a number of attractive features. It accords with our person-affecting sentiments. It naturally captures our asymmetric intuitions regarding the moral significance of creating future people. And it satisfies all four of Parfit's requirements: it addresses the Non-Identity Problem, it avoids the Repugnant and Absurd Conclusions, and it resolves the Mere-Addition Paradox. Furthermore, it fulfills two of these requirements in a particularly satisfying way: it avoids both the strong and weak versions of the Repugnant Conclusion, and it resolves the Mere-Addition Paradox in a way that preserves all of our initial judgments with respect to the three key cases.

In *Reasons and Persons*, Parfit (1984) posed a problem: provide a satisfying normative account that complies with four requirements. A number of people have suggested looking toward person-affecting views for a solution. The Saturated Harm Minimizing View vindicates this suggestion. It complies with Parfit's four requirements, and it offers an attractive solution to many of the problems in population ethics.⁴²

References

- Arrhenius, Gustaf. 2000. "An Impossibility Theorem for Welfarist Axiologies." *Economics and Philosophy* 16:247–266.
- Arrhenius, Gustaf. 2003. "The Person-Affecting Restriction, Comparativism, and the Moral Status of Potential People." *Ethical Perspectives* 10:185–195.

⁴²I'd like to thank Phil Bricker, Maya Eddon, Fred Feldman, Peter Graham, Elizabeth Harman, Julia Markovits, James Patten, Melinda Roberts, Ted Sider, Dennis Whitcomb, members of 2011 Bellingham Summer Philosophy Conference, and an anonymous referee, for helpful comments and discussion.

- Arrhenius, Gustaf. 2004. The Paradoxes of Future Generations and Normative Theory. In *The Repugnant Conclusion: Essays on Population Ethics*, ed. Jesper Ryberg and Torbjörn Tännsjö. Kluwer Academic Publishers pp. 201–218.
- Blackorby, Charles and David Donaldson. 1991. “Normative Population Theory: A Comment.” *Social Choice and Welfare* 8:261–267.
- Boonin-Vail, David. 1996. “Don’t Stop Thinking About Tomorrow: Two Paradoxes About Duties to Future Generations.” *Philosophy and Public Affairs* 25:267–307.
- Broome, John. 1992. *Counting the Costs of Global Warming*. White Horse Press.
- Broome, John. 1999. *Ethics out of Economics*. Cambridge University Press.
- Feldman, Fred. 1995. “Adjusting Utility for Justice: A Consequentialist Reply to the Objection from Justice.” *Philosophy and Phenomenological Research* 55:567–585.
- Hare, Caspar. 2007. “Voices From Another World: Must We Respect the Interests of People Who Do Not, and Will Never, Exist?” *Ethics* 117:498–523.
- Hare, R. 1993. Possible People. In *Essays in Bioethics*. Clarendon Press pp. 67–83.
- Holtug, Nils. 2004. Person-Affecting Moralities. In *The Repugnant Conclusion: Essays on Population Ethics*, ed. Jesper Ryberg and Torbjörn Tännsjö. Kluwer Academic Publishers.
- Huemer, Michael. 2008. “In Defense of Repugnance.” *Mind* 117.
- Lewis, David. 1986. *On The Plurality of Worlds*. Blackwell.
- Mackie, John L. 1985. The Parfit Population Problem. In *Persons and Values*. Clarendon Press.
- Narveson, Jan. 1967. “Utilitarianism and New Generations.” *Mind* 76:62–72.
- Ng, Yew-Kwang. 1989. “What Should We Do About Future Generations? Impossibility of Parfit’s Theory X.” *Economics and Philosophy* 5:235–253.
- Parfit, Derek. 1984. *Reasons and Persons*. Clarendon Press.
- Parsons, Josh. 2002. “Axiological Actualism.” *Australasian Journal of Philosophy* 80:137–147.
- Persson, Ingmar. 2004. The Root of the Repugnant Conclusion and its Rebuttal. In *The Repugnant Conclusion: Essays on Population Ethics*, ed. Jesper Ryberg and Torbjörn Tännsjö. Kluwer Academic Publishers.
- Rachels, Stuart. 1998. “Counterexamples to the Transitivity of Better Than.” *Australasian Journal of Philosophy* 76:71–83.

- Roberts, Melinda. 1998. *Child versus Childmaker: Future Persons and Present Duties in Ethics and the Law*. Rowman and Littlefield.
- Roberts, Melinda. 2003a. “Can it Ever Be Better Never to Have Existed At All? Person-Based Consequentialism and a New Repugnant Conclusion.” *Journal of Applied Philosophy* 20:159–185.
- Roberts, Melinda. 2003b. “Is the Person-Affecting Intuition Paradoxical?” *Theory and Decision* 55:1–44.
- Ryberg, Jesper. 1996. “Is the Repugnant Conclusion Repugnant?” *Philosophical Papers* 25:161–177.
- Ryberg, Jesper, Torbjörn Tännsjö and Gustaf Arrhenius. 2009. “The Repugnant Conclusion.” The Stanford Encyclopedia of Philosophy. URL = <http://plato.stanford.edu/archives/sum2009/entries/repugnant-conclusion/>.
- Sikora, Richard I. 1978. Is It Wrong to Prevent the Existence of Future Generations? In *Obligations to Future Generations*, ed. Richard I. Sikora and Brian Barry. The White Horse Press pp. 112–166.
- Tännsjö, Torbjörn. 2004. Why We Ought to Accept the Repugnant Conclusion. In *The Repugnant Conclusion: Essays on Population Ethics.*, ed. Jesper Ryberg and Torbjörn Tännsjö. Kluwer Academic Publishers.
- Temkin, Larry. 1987. “Intransitivity and the Mere Addition Paradox.” *Philosophy and Public Affairs* 16:138–187.
- Wolf, Clark. 1997. Person-Affecting Utilitarianism and Population Policy; Or, Sissy Jupe’s Theory of Social Choice. In *Contingent Future Persons*, ed. Nick Fotion and Jan C. Heller. Kluwer Academic Publishers pp. 99–122.
- Wrigley, Anthony. 2006. “Genetic Selections and Modal Harms.” *The Monist* 89:505–525.

Appendix

Proof: A Normative Theory Satisfies IIA_d iff it Meshes with IIA.

Here we’ll prove that a normative theory satisfies IIA_d iff it meshes with IIA.

Definitions: Let’s begin with the definitions required to make the meaning of this claim precise. I’ll say that the choice of an option A in a decision situation is *in accordance with* normative theory T iff T takes A to be permissible in that decision situation. I’ll say that the choice of an option A in a decision situation is *in accordance with* preference function f iff, for all available options X , $A \geq X$. And I’ll say that a set of choices S is *in accordance with* f/T iff all and only the choices in that set are in accordance with f/T . Finally, I’ll

say that a normative theory T meshes with preference constraint C iff the set S of choices that are in accordance with T is also in accordance with some preference function f that satisfies C .

We can characterize IIA and IIA_d as follows:

IIA: If there is a W_1W_2 -situation in which the W_1 -option \geq the W_2 -option, then in all W_1W_2 -situations the W_1 -option \geq the W_2 -option.

IIA_d :

- (i) If there exists a W_1W_2 -situation in which both the W_1 -option and the W_2 -option are permissible, then in all W_1W_2 -situations the W_1 -option is permissible iff the W_2 -option is permissible.
- (ii) If there exists a W_1W_2 -situation in which the W_1 -option is permissible and the W_2 -option is impermissible, then in all W_1W_2 -situations the W_2 -option is impermissible.

Proof: With this terminology in place, we can make sense of the result to be proved: a normative theory T satisfies IIA_d iff it meshes with IIA.

We'll prove the result in two parts. First (part I), we'll show that if a normative theory T violates IIA_d , then it will not mesh with IIA. Second (part II), we'll show that if a normative theory T satisfies IIA_d , then it will mesh with IIA. Together, these results entail the desired conclusion: that a normative theory T meshes with IIA iff it satisfies IIA_d .

Part I: If a normative theory T violates IIA_d , then it will not mesh with IIA.

We'll demonstrate this in two steps. First (I.A), we'll show that if a normative theory T violates the first clause of IIA_d , then the set of choices in accordance with T will only be in accordance with preference functions that violate IIA. Second (I.B), we'll show that if a normative theory T violates the second clause of IIA_d , then the set of choices in accordance with T will only be in accordance with preference functions that violate IIA.

I.A. The First Clause: First suppose a theory violates the first clause of IIA_d : there are W_1W_2 -situations in which both the W_1 -option and the W_2 -option are permissible according to T , and other W_1W_2 -situations in which one is permissible and the other not. Consider the set S of choices in accordance with T . Any preference function f in accordance with S must be such that, (i) in the W_1W_2 -situations in which both the W_1 -option and the W_2 -option are permissible, the W_1 -option \geq the W_2 -option and the W_2 -option \geq the W_a -option, and (ii) in W_1W_2 -situations in which (say) the W_1 -option is permissible and the W_2 -option is not, the W_2 -option $\not\geq$ the W_1 -option. This violates IIA.

I.B. The Second Clause: Suppose a theory violates the second clause of IIA_d : there are W_1W_2 -situations in which (say) the W_1 -option is permissible and the W_2 -option impermissible according to T , and other W_1W_2 -situations in which the W_2 -option is permissible. Consider the set S of choices in accordance with T . Any preference function f in accordance with S must be such that, (i) in the W_1W_2 -situations in which the W_1 -option is permissible and the W_2 -option impermissible, the W_1 -option \geq the W_2 -option and the W_2 -option $\not\geq$ the W_1 -option, and (ii) in the W_1W_2 -situations in which the W_2 option is permissible, the W_2 -option \geq the W_1 -option. This violates IIA.

Part II: If a normative theory T satisfies IIA_d , then the set of choices in accordance with T will mesh with IIA. (I.e., there will be some preference function f that's in accordance with this set of choices that meshes with IIA.)

Consider the preference functions f that are in accordance with the set S of choices that's in accordance with a normative theory T that satisfies IIA_d . Any such preference function f will either (i) mesh with IIA or (ii) not mesh with IIA. If f satisfies IIA, then we're done. If f doesn't satisfy IIA, then we'll show (II.A) that there's always a nearby preference function in accordance with S which *does* satisfy IIA. So no matter what, a comprehensive strategy in accordance with an IIA_d -satisfying theory T will be in accordance with some preference function f which satisfies IIA. So any normative theory T that satisfies IIA_d meshes with IIA.

II.A. The Key Result: Let S be the set of choices in accordance with a normative theory T that satisfies IIA_d , and let f be a preference function in accordance with S . If f violates IIA, then there is always a nearby preference function in accordance with S which *does* satisfy IIA.

Take any two W_1W_2 -situations in which f yields a violation of IIA with respect to its rankings of W_1 and W_2 in these situations. Since f violates IIA, it must be the case that the W_1 -option \geq the W_2 -option in one situation, and the W_1 -option $\not\geq$ the W_2 -option in the other. Call the first α and the second β .

Let's consider what set S of choices f could be in accordance with, given these constraints. In particular, let's consider the choices with respect to the W_1 and W_2 -options in α and β that f could be in accordance with. To start, we have 16 possibilities: in each situation, α and β , S could contain (i) the W_1 -option (and not the W_2 -option), (ii) the W_2 -option (and not the W_1 -option), (iii) both options or (iv) neither option.

Let's narrow this down.

First, S needs to be in accordance with a theory T that satisfies IIA_d . This rules out 6 possibilities, leaving us with 10 possibilities.⁴³

Second, S needs to be in accordance with f . Since f maintains at β that the W_1 -option $\not\geq$ the W_2 -option, it follows that S can't include the W_1 -option at β . This rules out 8 possibilities, 4 of which have already been ruled out, leaving us with 6 possibilities.⁴⁴ Likewise, since f maintains at α that the W_1 -option \geq the W_2 -option, it follows that S can't include the W_2 -option at α without also including the W_1 -option. This rules out 4 possibilities, 2 of which have already been ruled out, leaving us with 4 possibilities.⁴⁵

These are the four possible ways that S could treat the W_1 and W_2 -options in α and β that are compatible with the constraints we've imposed: (α : W_1 , β : neither), (α : neither, β : W_2), (α : both, β : neither), (α : neither, β : neither).

Now consider two preference functions, f_1 and f_2 , which are the same as f in every respect except for their preference rankings of the W_1 and W_2 -options in α and β . While f maintains that the W_1 -option \geq the W_2 -option in α and the W_1 -option $\not\geq$ the W_2 -option

⁴³The 6 possibilities this rules out are: (α : W_1 , β : W_2), (α : W_1 , β : both), (α : W_2 , β : W_1), (α : W_2 , β : both), (α : both, β : W_1), (α : both, β : W_2).

⁴⁴The 4 additional possibilities this rules out are: (α : W_1 , β : W_1), (α : both, β : W_1), (α : both, β : both), (α : neither, β : both).

⁴⁵The 2 additional possibilities this rules out are: (α : W_2 , β : W_1), (α : W_2 , β : both).

in β , f_1 maintains that the W_1 -option \geq the W_2 -option in both, and f_2 maintains that the W_1 -option $\not\geq$ the W_2 -option in both. In each of the four possibilities for S compatible with the constraints, either f_1 or f_2 will be in accordance with S . (f_1 is in accordance with $(\alpha : W_1, \beta : \text{neither})$, f_2 is in accordance with $(\alpha : \text{neither}, \beta : W_2)$, and both are in accordance with $(\alpha : \text{both}, \beta : \text{neither})$ and $(\alpha : \text{neither}, \beta : \text{neither})$.) And both f_1 and f_2 are compatible with IIA with respect to the W_1 and W_2 -options in α and β .

These nearby preference functions only ‘fix’ f with respect to one violation of IIA. But by iterating this process, we can transform any f in accordance with S which fails to satisfy IIA into a nearby alternative which is also in accordance with S and which does satisfy IIA.

Proof: Given IIA_d and that Some Option is Permissible, the Three Judgments Yield a Contradiction.

Here we’ll see that given IIA_d and the assumption that some option is always permissible, our intuitive judgments in the three cases that comprise the Mere-Addition Paradox lead to a contradiction.

The intuitive judgments that are reported with respect to these three cases leave a bit of wiggle room. It is usually left open in the first case whether both options are intuitively permissible or whether only the W_2 -option is permissible. Likewise, it is usually left open in the third case whether both options are intuitively permissible or whether only the W_1 -option is permissible. This gives us four permutations. We’ll show that all four of these possible prescriptions lead to contradictions.

First, consider the most natural judgments: suppose that both options are permissible in case one, and that only the W_1 -option is permissible in case three. And consider the case in which the agent has a choice between all three of the outcomes:

W_1	W_2		W_3	
$a1 - a10$	$b1 - b10$	$b11 - b20$	$c1 - c10$	$c11 - c20$
+10	+10	+5	+9	+9

Given the judgment in the first case, IIA_d entails that in W_1W_2 -situations the W_1 -option is permissible *iff* the W_2 -option is permissible. Given the second judgment, IIA_d entails that in W_2W_3 -situations the W_2 -option is impermissible. It follows that in $W_1W_2W_3$ -situations, both the W_1 and W_2 -options are impermissible. Given the third judgment, IIA_d entails that in W_1W_3 -situations the W_3 -option is impermissible. It follows that in $W_1W_2W_3$ -situations like this one, all three options are impermissible. But there must always be a permissible option available. Contradiction.

Second, suppose that only the W_2 -option is permissible in case one, and only the W_1 -option is permissible in case three. Then this will change what the first judgment and IIA_d entail in the initial case: they will now entail that in W_1W_2 -situations (and *a fortiori* $W_1W_2W_3$ -options) the W_1 -option is impermissible. Since the second and third judgments and IIA_d entail that the W_2 and W_3 -options are also impermissible in these situations, we again get the result that all three options are impermissible. But there must be a permissible option. Contradiction.

Third, suppose that both options are permissible in both cases one and three. Then this will change what the third judgment and IIA_d entail in the initial case: they will now entail that in $W_1W_2W_3$ -situations, the W_1 -option is permissible *iff* the W_3 -option is permissible. Since the first judgment and IIA_d entail that the W_1 -option is permissible *iff* the W_2 -option is permissible in these situations, it follows that all three options are either permissible or impermissible. And since the second judgment and IIA_d entail that the W_2 -option is impermissible in these situations, we get the result that all three options are impermissible. But there must be a permissible option. Contradiction.

Fourth, suppose that only the W_2 -option is permissible in case one, and both options are permissible in case three. Then in $W_1W_2W_3$ -situations, the first judgment and IIA_d will entail that the W_1 -option is impermissible, the second judgment and IIA_d will entail that the W_2 -option is impermissible, and the third judgment and IIA_d will entail that the W_3 -option is permissible *iff* the W_1 -option is permissible. Together this entails that all three options are impermissible. But there must be a permissible option. Contradiction.