Intrapersonal Arguments for the Repugnant Conclusion

Tomi Francis

June 22, 2023

Abstract

In "An Intrapersonal Addition Paradox", Jacob Nebel provides a novel "intrapersonal" argument for the Repugnant Conclusion. The most controversial premise of Nebel's argument is the "Probable Addition Principle", on which it is better for individuals to receive additional chances of existence with a life worth living. I provide an alternative intrapersonal argument for the Repugnant Conclusion which does not assume the Probable Addition Principle. I also show that Pareto principles alone, when conjoined with very minimal principles of prudence, imply a version of the Repugnant Conclusion.*

1 Introduction

Most people find the following proposition very difficult to believe:

> *The 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.[1]

However, there are powerful arguments for the Repugnant Conclusion; consequently, avoiding it may be even more difficult than believing it. The most influential such argument is Parfit's Mere Addition Paradox, but there are many others.[2] Most of these are "interpersonal" arguments, which appeal directly to plausible principles for comparing populations in terms of their overall value ("interpersonal principles").

Jacob Nebel has recently pioneered another kind of argument for the Repugnant Conclusion: an *intrapersonal* argument, which he calls the Intrapersonal Addition Paradox.[3] Rather than appealing to interpersonal principles, intrapersonal arguments appeal to principles governing which prospects are better for *individuals* ("intrapersonal principles"). They use these intrapersonal principles, together with certain bridge principles, to derive betterness comparisons of outcomes or prospects.

This paper is about intrapersonal arguments for the Repugnant Conclusion, with particular attention to Nebel's Intrapersonal Addition Paradox. I recap this argument in §2. The Intrapersonal Addition Paradox is a powerful argument, but it has one particularly controversial premise: the "Probable Addition Principle". This principle says (roughly) that it is better for a person to receive additional chances of existence with a good life. It is controversial in large part because it seems hard to square with

> *Non-Comparativism* An outcome in which S exists cannot be better or worse for S than an outcome in which S does not exist.

In §3, I make more precise the sense in which Probable Addition is incompatible with Non-Comparativism: the two principles are inconsistent if we assume a version of the Sure Thing Principle. It is not completely obvious that Non-Comparativists should accept the Sure Thing Principle.

However, I show that without it, Nebel's positive argument for the Probable Addition Principle does not succeed. Either way, Non-Comparativists have little reason to accept the Probable Addition Principle, and thus little reason to worry about the Intrapersonal Addition Paradox, insofar as Non-Comparativism is defensible.

In §4, however, I provide a new intrapersonal argument for the Repugnant Conclusion which replaces the Probable Addition Principle with the "Conditional Value Principle", which says that if $\mathcal{X}$ gives S a non-zero probability of existence with a good life, otherwise non-existence, and $\mathcal{Y}$ gives S a non-zero probability of existence with a bad life, otherwise non-existence, then $\mathcal{X}$ is better for S than $\mathcal{Y}$. The Conditional Value Principle is intuitively much more compelling than the Probable Addition Principle, and does not tacitly assume Comparativism. I argue that in light of this new intrapersonal argument, the only plausible response for opponents of the Repugnant Conclusion is to reject "Weak Pareto for Equal Risk", which is the ex ante Pareto principle which appears in the Intrapersonal Addition Paradox, as well as in my argument from the Conditional Value Principle. In §5, I further substantiate this last claim by showing that, given one very minimal intrapersonal principle, Weak Pareto for Equal Risk alone has an implication which seems to me about as "repugnant" as the Repugnant Conclusion. Incidentally, unlike most arguments for repugnant conclusions, my argument in §5 does not assume the transitivity (or acyclicity) of the betterness relation. I conclude in §6 with some thoughts on the choice between Weak Pareto for Equal Risk and avoidance of the Repugnant Conclusion.

2 The Intrapersonal Addition Paradox

The Intrapersonal Addition Paradox can be broken down into two stages. The first stage, which Nebel calls the "probable addition argument", aims to establish the

> *Intrapersonal Repugnant Conclusion* For any person S, there is some probability p such that any prospect in which S would have a wonderful life with probability p or less, and would otherwise never exist, is worse for S than a certainly mediocre life.[4]

Some of these terms require explanation. By a "mediocre" life (or wellbeing level), I mean one that is barely worth living: one that is better than a neutral life, but only slightly.[5] By a "wonderful" or "excellent" life (or wellbeing level), I mean one that contains large quantities of whatever it is that makes life worth living, and very little of what makes life worse.

The second "interpersonal" stage of the argument derives the Repugnant Conclusion from the Intrapersonal Repugnant Conclusion, via four bridge principles. We now turn to the probable addition argument.

2.1 The Probable Addition Argument

The probable addition argument has two premises. The first is the Probable Addition Principle. This principle says that in cases where existence is not guaranteed, it is always better for an individual to get more wellbeing in the states of nature in which she would have existed anyway, as well as (perhaps) getting a life worth living in some states of nature in which she would not have existed otherwise. Formally, it says:

> *Probable Addition Principle* For any prospects $\mathcal{X}$ and $\mathcal{Y}$ and any person S who might exist in those prospects: if, in every state of nature in which S would exist in Y, S would be better off in X, and if, in every other state of nature, S's life would be worth living in X, X is better for S than Y.[6]

We shall discuss the Probable Addition Principle at length in §3, but it is worth noting at the outset that the Probable Addition Principle, while hardly implausible, is not all that compelling

either. Rather than getting a small chance of existence with an excellent life and non-existence otherwise, is it better to be slightly better off in that state of nature, and get a life that's barely worth living otherwise? It may well be, but it's not outlandish to think that it isn't. Perhaps it matters that the first prospect guarantees that one has an excellent life, conditional on existence, while the second all but guarantees that one has a mediocre life, conditional on existence. More generally, while it's clearly a good thing to be better off in some state of nature, it's far less clear that it's a good thing to get additional chances of existence with a good life; and it's not obviously a mistake to think that the combination of the two might fail to be for the better, all things considered.

The second premise says that it is better for an individual to get a small improvement to their wellbeing level for sure, rather than getting a sufficiently tiny chance of a much larger improvement to their wellbeing level. Or, more precisely:

> *Minimal Prudence* For any individual S and very high welfare level a, there are some mediocre welfare levels z and z^- (where $z \succ z^-$) and some probability p such that any prospect in which S is certain to exist at level z is better for S than any prospect in which S might, with any probability less than or equal to p, exist at level a, and would otherwise exist at level z^-.[7]

I share Nebel's opinion that judgements like this are beyond serious doubt, and so I shall assume that rejecting Minimal Prudence is off the table.[8]

The probable addition argument proceeds as follows. Let $a^+ \succ a$ be excellent wellbeing levels, and let $z \succ z^-$ be mediocre wellbeing levels.[9] By Minimal Prudence, there is some probability p such that certainty of existence at z is better than probability p of existence at a^+ and otherwise existence at z^-. Consider now the following three prospects $\mathcal{A}$, $\mathcal{A}^+$ and $\mathcal{Z}$, where the wellbeing levels specified are the wellbeing levels of some person S, and a dash represents the non-existence of S:

Table 1

	State 1 (probability p)	State 2 (probability $1-p$)
$\mathcal{A}$	a	-
$\mathcal{A}^+$	a^+	z^-
$\mathcal{Z}$	z	z

The Probable Addition Principle implies that $\mathcal{A}^+$ is better for S than $\mathcal{A}$. Minimal Prudence implies that $\mathcal{Z}$ is better for S than $\mathcal{A}^+$. Transitivity of the at-least-as-good-as relation then implies that $\mathcal{Z}$ is better for S than $\mathcal{A}$, which is the Intrapersonal Repugnant Conclusion.[10]

Next, we shall use the Intrapersonal Repugnant Conclusion to derive the (interpersonal) Repugnant Conclusion.

2.2 The Interpersonal Stage

The interpersonal stage of the Intrapersonal Addition Paradox has four more premises. First, some terminology: a prospect is *egalitarian* if and only if each person who might exist has the same chances of existing at each wellbeing level as everyone else (so that there is perfect ex ante equality), and each possible outcome is a perfectly equal population (so that there is perfect ex post equality). We can now state the four premises.

Stochastic Indifference for Equal Risk says that if two prospects are egalitarian and equally good in each state of nature, then they are equally good overall:

> *Stochastic Indifference for Equal Risk* For any egalitarian prospects $\mathcal{X}$ and $\mathcal{Y}$, if every possible outcome of $\mathcal{X}$ and every possible outcome of $\mathcal{Y}$ are equally good, then $\mathcal{X}$ and $\mathcal{Y}$ are equally good.[11]

Weak Pareto for Equal Risk says that if one egalitarian prospect is better for everyone than another, it is better overall:

> *Weak Pareto for Equal Risk* For any egalitarian prospects $\mathcal{X}$ and $\mathcal{Y}$, if $\mathcal{X}$ is better than $\mathcal{Y}$ for each person who might exist in either prospect, then $\mathcal{X}$ is better than $\mathcal{Y}$.[12]

The *Same-Number Quality Claim* says that populations involving the same numbers of people, all at the same level of wellbeing, must be equally good overall:

> *Same-Number Quality Claim* Any two populations consisting of the same number of people, all at the same levels of well-being, are equally good.[13]

Finally, *Certainty Equivalence* says that the betterness relation on populations is the same as the betterness relation on the corresponding riskless prospects:

> *Certainty Equivalence* For any riskless prospects $\mathcal{X}$ and $\mathcal{Y}$, which guarantee populations X and Y respectively, $\mathcal{X}$ is better than $\mathcal{Y}$ just in case X is better than Y.[14]

The interpersonal stage of the argument goes as follows. Let A be any population of m lives at excellent wellbeing level a. Let p be a sufficiently small probability that certainty of life at the mediocre level z is better than a p chance of life at level a and non-existence otherwise (by the Intrapersonal Repugnant Conclusion). Given this small probability p, we can construct a set of states of nature $s_1, \ldots s_n$, where each state has probability $1/n < p$. We can also construct n disjoint sets of possible people $G_1, G_2, \ldots, G_n$, each containing m people, and with G_1 being the set of A-people; write G to denote the set of all people who belong to some G_i. Some notation: if X is any set of people, we write $X[w]$ to denote the outcome in which the X people exist at wellbeing level w, and nobody else exists. Now consider the following three prospects:

Table 2

	s_1	s_2	$\ldots$	s_n
$\mathcal{A}$	$G_1[a]$	$G_1[a]$	$\ldots$	$G_1[a]$
$\mathcal{A}'$	$G_1[a]$	$G_2[a]$	$\ldots$	$G_n[a]$
$\mathcal{Z}$	$G[z]$	$G[z]$	$\ldots$	$G[z]$

Notice that all three prospects are egalitarian. The Same-Number Quality Claim and Stochastic Indifference for Equal Risk thus imply that $\mathcal{A}$ and $\mathcal{A}'$ are equally good. Since certainty of z is better for each G-person than a tiny probability of a and non-existence otherwise (by the Intrapersonal Repugnant Conclusion), Weak Pareto for Equal Risk implies that $\mathcal{Z}$ is better than $\mathcal{A}'$. Transitivity then implies that $\mathcal{Z}$ is better than $\mathcal{A}$. Since $\mathcal{A}$ and $\mathcal{Z}$ guarantee populations A and $G[z]$ respectively, Certainty Equivalence implies that $G[z]$ is better than A, which is the Repugnant Conclusion.

3 The Probable Addition Principle

For most of us, accepting the Repugnant Conclusion will be a last resort.[15] To avoid this last resort, we need to deny at least one of the six premises of the Intrapersonal Addition Paradox. Some of these premises, I think, are not up for rejection. Minimal Prudence is one of them. It is also not promising to reject Certainty Equivalence, Stochastic Indifference for Equal Risk or the Same-Number Quality Claim. Even if we deny all three of these principles, we would still be left with the conclusion that $\mathcal{Z}$ is better than $\mathcal{A}'$, which seems about as "repugnant" as the Repugnant Conclusion.

Those who wish to robustly avoid the Repugnant Conclusion thus need to deny either the Probable Addition Principle or Weak Pareto for Equal Risk. On the face of it, the obvious choice here is

to reject the Probable Addition Principle: while Weak Pareto for Equal Risk encodes the eminently plausible idea that overall betterness should align with betterness for individuals when there is no risk of inequality, the Probable Addition Principle is more dubious. This is mainly because the Probable Addition Principle is in tension with another plausible view: *Non-Comparativism*.[16] This is the view that an outcome in which S exists cannot be better or worse for S than an outcome in which S does not exist.

Technically, Non-Comparativism is outright inconsistent with the Probable Addition Principle.[17] To see this, note that if S certainly exists with a good life in $\mathcal{X}$, and certainly does not exist in $\mathcal{Y}$, then S is better off in $\mathcal{X}$ for all states of nature in which she would exist in $\mathcal{Y}$ (because there are none), and has a life worth living in $\mathcal{X}$ for all other states of nature. The Probable Addition Principle therefore implies that $\mathcal{X}$ is better for S than $\mathcal{Y}$, contradicting Non-Comparativism.[18]

There is an easy reply to this argument. The Probable Addition Principle might be intended to apply only when there is some state of nature in which $\mathcal{X}$ makes S better off than she is in $\mathcal{Y}$. In that case, assuming Non-Comparativism, S must have a non-zero probability of existence in $\mathcal{X}$, which blocks the preceding argument.

Still, while this weaker version of the Probable Addition Principle is not immediately inconsistent with Non-Comparativism, the two principles do not sit well with each other. Suppose we also assume an evaluative version of the Sure Thing Principle, which says that when we compare two prospects, we can ignore states of nature which yield the same outcome. Or, more precisely, it says:

> *Sure Thing Principle*[19] If prospects $\mathcal{X}$ and $\mathcal{Y}$ yield the same outcome in state s_i, then $\mathcal{X}$ is at least as good for S as $\mathcal{Y}$ if and only if $\mathcal{X}$ is at least as good for S as $\mathcal{Y}$ conditional on the non-occurrence of state s_i.

To see the tension between the Probable Addition Principle and Non-Comparativism, let $a \succ b$ be wellbeing levels corresponding to lives worth living, and consider the following three prospects for person S over two equi-probable states of nature:

Table 3

	State 1	State 2
$\mathcal{P}_1$	b	-
$\mathcal{P}_2$	a	b
$\mathcal{P}_3$	b	a

Intuitively, $\mathcal{P}_2$ and $\mathcal{P}_3$ are equally good, since they guarantee existence and give the same chances of the same wellbeing levels. The Probable Addition Principle implies that $\mathcal{P}_2$ is better than $\mathcal{P}_1$. Transitivity then implies that $\mathcal{P}_3$ is better than $\mathcal{P}_1$. Applying the Sure Thing Principle to ignore State 1 for this last comparison, we find that certain existence at wellbeing level a is better for S than certain non-existence, which violates Non-Comparativism (given Certainty Equivalence). Non-Comparativists who accept the Sure Thing Principle should therefore reject the Probable Addition Principle.

One might doubt that Non-Comparativists should accept the application of the Sure Thing Principle to cases involving risks of non-existence.[20] Without it, however, the positive argument Nebel provides for the Probable Addition Principle is not persuasive.[21] Suppose that a is an excellent level of wellbeing, d is some small additional quantity of wellbeing, b is a wellbeing level that is slightly bad (that is, slightly worse than the level of a neutral life), y is a mediocre wellbeing level, and p and q are probabilities. Nebel's argument for the Probable Addition Principle appeals to the following prospects for S:

Table 4

	State 1 $(1-p)(1-q)$	State 2 $(p(1-q))$	State 3 (q)
$\mathcal{A}$	a	a	-
$\mathcal{A}'$	$a+d$	b	-
$\mathcal{A}^+$	$a+d$	$a+d$	y

Nebel notes that if we ignore State 3 (i.e., set $q = 0$), then if p is sufficiently small, $\mathcal{A}'$ is surely better for S than $\mathcal{A}$.[22] Since these prospects have the same outcome in State 3, he claims that $\mathcal{A}'$ must be better than $\mathcal{A}$ even if $q \neq 0$. Similarly, $\mathcal{A}^+$ is clearly better than $\mathcal{A}'$ if we ignore State 1; since these prospects give the same result in State 1, it seems we can conclude that $\mathcal{A}^+$ is better than $\mathcal{A}'$. Transitivity then implies that $\mathcal{A}^+$ is better than $\mathcal{A}$. Since d, q and y were chosen arbitrarily, this is sufficient to establish the Probable Addition Principle.

However, the claim that $\mathcal{A}^+$ is better than $\mathcal{A}'$ is only compelling if we ignore State 3 (thus tacitly assuming the Sure Thing Principle). To see this, consider the following prospects, where the numbers represent years of good life:

Table 5

	State 1 (0.01)	State 2 (0.00001)	State 3 (0.98999)
$\mathcal{A}'$	101	-1	-
$\mathcal{A}^+$	101	101	1

$\mathcal{A}^+$ is not *clearly* better than $\mathcal{A}'$, because it involves moving from a situation in which one is almost certain that one will enjoy an excellent life if one exists to a situation in which one is almost certain to have a mediocre life if one exists. It's not crazy to think that this could make $\mathcal{A}^+$ worse than $\mathcal{A}'$. The claim that $\mathcal{A}^+$ is better than $\mathcal{A}'$ is, at best, only marginally more compelling than the Probable Addition Principle itself.

Here's the upshot. If we do apply the Sure Thing Principle to risky existence cases, the Probable Addition Principle is incompatible with Non-Comparativism. If we do not, then Nebel's positive argument for the Probable Addition Principle is not compelling. The Probable Addition Principle is therefore a shaky foundation for the Intrapersonal Addition Paradox. However, it turns out that we can do without it; showing this properly is the main point of this paper. In the next section, I provide an intrapersonal argument for the Repugnant Conclusion which does not assume the Probable Addition Principle.

4 Repugnance Without Probable Addition

For this argument, we shall need all of the premises of the Intrapersonal Addition Paradox apart from the Probable Addition Principle, which we shall replace with a more compelling principle. To help state it, let me introduce some notation. Let us say that a prospect is *conditionally good (bad)* for S if it guarantees S a good (bad) life, provided she exists, and gives her a non-zero probability of existence. The principle we shall use in place of the Probable Addition Principle is the

> *Conditional Value Principle* If $\mathcal{X}$ is conditionally good for S and $\mathcal{Y}$ is conditionally bad for S, then $\mathcal{X}$ is better for S than $\mathcal{Y}$.[23]

The Conditional Value Principle can be supported by the following brief argument. A conditionally good prospect is good for the person receiving it. A conditionally bad prospect, on the other hand, is bad for the person receiving it. If something is good for a person, and another thing is bad for

the same person, the first thing is better for her than the second. The Conditional Value Principle follows from these three claims.

Of course, this argument only works if it makes sense to say that prospects involving risks of non-existence can be good or bad for a person. Comparativists should have no trouble here: they can say that such prospects are good (or bad) for a person if and only if they are better (or worse) than certain non-existence. Non-Comparativists cannot say this. But some will agree with Parfit that, even if it cannot be better or worse for a person to exist than to not exist, it may yet be non-comparatively good or bad for a person to exist.[24] It seems natural to extend this concept of non-comparative value to the case of prospects involving risks of non-existence. Non-Comparativists who disagree with Parfit will reject my argument, but may still find the Conditional Value Principle independently compelling.

Next, we shall need to strengthen two of the other premises, but in ways which do not make much difference to their intuitive plausibility. First, we shall need to assume a stronger version of Minimal Prudence, which applies in cases in which a person is not guaranteed to exist. This strengthened principle roughly says that for any wellbeing levels $a \succ b$ and $c \succ d$, it is better for S to get a rather than b in some state of nature, even if this comes at the cost of getting d rather than c in another state of nature with some much smaller probability. Moreover, this holds even if S might not exist.

To state this principle more formally, it will be helpful to introduce some more notation. If $w_1, w_2, \ldots, w_n$ are wellbeing levels and $p_1, p_2 \ldots p_n$ are probabilities adding up to 1, we write $(w_1[p_1], \ldots, w_n[p_n])$ to denote any prospect for S which gives w_1 with probability p_1, w_2 with probability p_2, and so on. For this notation, we shall denote non-existence by Ω. We can now state our strengthening of Minimal Prudence:

> *Prudence* Let $a \succ b$ and $c \succ d$ be any wellbeing levels. There exists some sufficiently small probability p' such that for any $p \leq p'$ and any probability of non-existence q,
>
> $$(a[p(1-q)], d[(1-p)(1-q)], \Omega[q])$$
>
> is better for S than
>
> $$(b[p(1-q)], c[(1-p)(1-q)], \Omega[q])$$
>
> provided S exists in exactly the same states of nature in each case.

Prudence is supported by the same sorts of considerations which favour Minimal Prudence: decision theories which do not satisfy Prudence are reckless in the sense that they sometimes prioritise what happens in states of nature with arbitrarily tiny probabilities over what happens in states with proportionally much larger probabilities.

The second principle we shall need to strengthen is Stochastic Indifference for Equal Risk. We shall assume the slightly stronger

> *Statewise Dominance for Equal Risk* For any egalitarian prospects $\mathcal{X}$ and $\mathcal{Y}$ over the same states of nature, if each possible outcome of $\mathcal{X}$ is at least as good as its same-state counterpart in $\mathcal{Y}$, then $\mathcal{X}$ is at least as good as $\mathcal{Y}$. If, additionally, some outcome of $\mathcal{X}$ is better than the corresponding outcome of $\mathcal{Y}$, $\mathcal{X}$ is better than $\mathcal{Y}$.

Statewise Dominance for Equal Risk is supported by the same considerations which lend credence to Stochastic Indifference for Equal Risk: if a prospect has some chance of being better, and is certain to be at least as good, the prospect is better (at least when perfect ex post and ex ante equality is guaranteed).

We now have all the premises we need. Before beginning the argument proper, we first need to derive the

> *Absolute Value Principle*[25] Let a be any wellbeing level corresponding to a life worth living, and b any level of a life worth not living. If population X consists solely of lives at a, and Y consists solely of lives at b, X is better than Y.

Let n and m be any numbers of people, a any good wellbeing level, and b any bad wellbeing level. Let $G_1, \ldots G_m$ be disjoint sets of n people each, and let $H_1, \ldots, H_n$ be disjoint sets of m people each, such that the G_i and the H_j sets contain the same $n \cdot m$ people. (That is, $\bigcup_{i=1}^{m} G_i = \bigcup_{i=1}^{n} H_i$.) Consider $n \cdot m$ equi-probable states of nature, which can be partitioned into events in two ways: (i) $s_1, \ldots, s_m$, where each s_i is the disjunction of n states of nature; (ii) $T_1, \ldots, T_n$, where each T_k is the disjunction of m states of nature. Now consider the following two prospects illustrated by Tables 6 and 7 below:

Table 6

	s_1	s_2	$\ldots$	s_m
$\mathcal{A}$	$G_1[a]$	$G_2[a]$	$\ldots$	$G_m[a]$

Table 7

	T_1	T_2	$\ldots$	T_n
$\mathcal{B}$	$H_1[b]$	$H_2[b]$	$\ldots$	$H_n[b]$

The Conditional Value Principle implies that $\mathcal{A}$ is better than $\mathcal{B}$ for each person who might exist; Weak Pareto for Equal Risk then implies that $\mathcal{A}$ is better than $\mathcal{B}$. Note that $\mathcal{A}$ guarantees the existence of n people at level a, while $\mathcal{B}$ guarantees the existence of m people at level b. The Same-Number Quality Claim, Stochastic Indifference for Equal Risk and Certainty Equivalence therefore imply that an arbitrary population of n people at level a must be better than an arbitrary population of m people at level b; since a, b, n and m were chosen arbitrarily, we can conclude that the Absolute Value Principle is true.

We will not need the Conditional Value Principle from this point on: its only purpose was to provide an intrapersonal justification for the Absolute Value Principle. (That said, the Absolute Value Principle is intuitively compelling in its own right.) Next, we shall derive the Repugnant Conclusion from Weak Pareto for Equal Risk, Statewise Dominance for Equal Risk, Prudence, the Same-Number Quality Claim, Certainty Equivalence, and the Absolute Value Principle.

Let A be any population of m lives at some excellent level of wellbeing a. Let $a^+ \succ a$ be a slightly higher level of wellbeing. Let $z \succ z^-$ be mediocre wellbeing levels, and let b be the level of a slightly bad life. Naturally, we have $z^- \succ b$ and $a \succ z$.

Applying Prudence twice and taking the minimum of the two probabilities generated by the two applications of this principle, we find that there is some small probability p' such that for any $p < p'$ and any probability of non-existence q:

(i)

$$(b[p(1-q)], a^+[(1-p)(1-q)], q[\Omega])$$

is better for S than

$$(a[1-q], \Omega[q])$$

and

(ii)

$$(b[p(1-q)], z[(1-p)(1-q)], \Omega[q])$$

is better for S than

$$(a^+[p(1-q)], z^-[(1-p)(1-q)], \Omega[q])$$

That is, a sufficiently small chance of getting b but larger chance of getting a^+ is better than getting a for sure (conditional on existence in the same states of nature), and a sufficiently small chance of getting b but larger chance of getting z is better than the same small chance of getting a^+ and larger chance of getting z^- (conditional on existence in the same states of nature). Now define n to be the smallest positive integer such that $1/n < p'/2$, and write $p = 1/n$.

We will consider prospects over states of nature $s_1, \ldots, s_{n^2}, T_1, \ldots, T_{n^2}$, where each s_i has probability $p^2(1-p)$ and each T_i has probability p^3. (These probabilities add up to 1.) We can construct disjoint sets $G_1, \ldots, G_{n^2}$, each containing m possible people, and where G_1 contains the A-people. Now consider the following two prospects:

Table 8

	s_1	...	s_{n^2}	T_1	...	T_{n^2}
$\mathcal{R}_1$	$G_1[a]$	...	$G_1[a]$	$G_1[a]$	...	$G_1[a]$
$\mathcal{R}_2$	$G_1[a]$	...	$G_{n^2}[a]$	$G_1[a]$	...	$G_{n^2}[a]$

Since $\mathcal{R}_1$ and $\mathcal{R}_2$ are both egalitarian prospects, and both guarantee the existence of m people at wellbeing level a, the Same-Number Quality Claim and Statewise Dominance for Equal Risk imply that these prospects are equally good. Next, consider

Table 9

	s_1	...	s_{n^2}	T_1	...	T_{n^2}
$\mathcal{R}_2$	$G_1[a]$	...	$G_{n^2}[a]$	$G_1[a]$	...	$G_{n^2}[a]$
$\mathcal{R}_3$	$G_1[a^+]$	...	$G_{n^2}[a^+]$	$G_1[b]$	...	$G_{n^2}[b]$

$\mathcal{R}_3$ is an egalitarian prospect, and each person exists in precisely the same states of nature in each prospect. Prudence therefore implies that $\mathcal{R}_3$ is better than $\mathcal{R}_2$ for every person who might exist, since each person is less than $2p$ times as likely to receive b as they are to receive a^+. Weak Pareto for Equal Risk then implies that $\mathcal{R}_3$ is better than $\mathcal{R}_2$. Next, we have

Table 10

	s_1	...	s_{n^2}	T_1	...	T_{n^2}
$\mathcal{R}_3$	$G_1[a^+]$	...	$G_{n^2}[a^+]$	$G_1[b]$	...	$G_{n^2}[b]$
$\mathcal{R}_4$	$G_1[a^+]$	...	$G_{n^2}[a^+]$	$G[z^-]$	...	$G[z^-]$

Once again, $\mathcal{R}_4$ is egalitarian. The two prospects are equally good in states s_1 to s_{n^2} (because they yield the same outcomes), and the Absolute Value Principle implies that $\mathcal{R}_4$ is better than $\mathcal{R}_3$ in states T_1 to T_{n^2}. Statewise Dominance for Equal Risk therefore implies that $\mathcal{R}_4$ is better than $\mathcal{R}_3$. Next, consider

Table 11

	s_1	...	s_{n^2}	T_1	...	T_{n^2}
$\mathcal{R}_4$	$G_1[a^+]$	...	$G_{n^2}[a^+]$	$G[z^-]$	...	$G[z^-]$
$\mathcal{R}_5$	$G_1[b]$	...	$G_{n^2}[b]$	$G[z]$	...	$G[z]$

$\mathcal{R}_5$ is egalitarian, and each person exists in precisely the same states of nature in each prospect. $\mathcal{R}_4$ gives each person probability $p^2(1-p)$ of getting a^+ and probability p of getting z^-. Thus, the probability of getting a^+ rather than b is less than $2p$ times as much as the probability of getting

z^- rather than z. Prudence therefore implies that $\mathcal{R}_5$ is better than $\mathcal{R}_4$ for each person who might exist; Weak Pareto for Equal Risk then implies that $\mathcal{R}_5$ is better than $\mathcal{R}_4$. Finally, consider

Table 12

	s_1	...	s_{n^2}	T_1	...	T_{n^2}
$\mathcal{R}_5$	$G_1[b]$	...	$G_{n^2}[b]$	$G[z]$	...	$G[z]$
$\mathcal{R}_6$	$G[z]$	...	$G[z]$	$G[z]$	...	$G[z]$

$\mathcal{R}_6$ is egalitarian. As in the case of $\mathcal{R}_4$ and $\mathcal{R}_3$, $\mathcal{R}_5$ and $\mathcal{R}_6$ have the same (and therefore equally good) outcomes in each state T_i, and the Absolute Value Principle implies that $\mathcal{R}_6$ is better than $\mathcal{R}_5$ in each state s_i. Statewise Dominance for Equal Risk therefore implies that $\mathcal{R}_6$ is better than $\mathcal{R}_5$. Putting all of our claims together with transitivity, we have that $\mathcal{R}_6$ is better than $\mathcal{R}_1$; Certainty Equivalence then implies that $G[z]$ is better than A, which is the Repugnant Conclusion.

The point of this argument is to show that intrapersonal arguments for the Repugnant Conclusion can go through without the Probable Addition Principle. Prudence is only slightly less compelling than Minimal Prudence, while Statewise Dominance for Equal Risk is just as compelling as Stochastic Indifference for Equal Risk. The Conditional Value Principle seems to me a bare minimum requirement on any theory of prudence which countenances comparisons of prospects which involve different probabilities of existence. Thus, unless we are prepared to throw out different-probability comparisons altogether, our only serious option for blocking the argument from the Conditional Value Principle is to reject Weak Pareto for Equal Risk.[26] To further substantiate this claim, I shall next show that, when supplemented with just one very minimal intrapersonal principle, Weak Pareto for Equal Risk implies an intuitively "repugnant" conclusion.

5 Pareto Principles and Repugnant Conclusions

The intrapersonal principle in question says that if S has the same chances of getting each wellbeing level between two prospects, these prospects are equally good for her. It also says that if S has better chances of being better off in one prospect, then that prospect is better for S. Or, more formally:

> *Stochastic Dominance for Personal Prospects* Suppose that S exists in precisely the same states of nature in prospects $\mathcal{X}$ and $\mathcal{Y}$. If, for each wellbeing level w, $\mathcal{Z}$ gives S at least as great a probability of getting at least w as $\mathcal{Y}$ does, then $\mathcal{X}$ is better than $\mathcal{Y}$. If, additionally, there is some wellbeing level w for which $\mathcal{X}$ gives S a greater chance of getting at least w than $\mathcal{Y}$ does, then $\mathcal{X}$ is better for S than $\mathcal{Y}$.

Note that because this principle only applies when S exists in the same states of nature in both prospects, it is compatible with the combination of Non-Comparativism and "deference" based responses to the opaque sweetening problem discussed by Caspar Hare.[27]

To see how Stochastic Dominance for Personal Prospects spells trouble when conjoined with Weak Pareto for Equal Risk, consider the following two prospects over equi-probable states of nature s_1 to s_{n+1}, where G_1 to G_n are disjoint sets of m people each (where m is any number), G is the set of all people in some G_i, a is an excellent wellbeing level, and $z \succ z^-$ are mediocre wellbeing levels:

Table 13

	s_1	...	s_n	s_{n+1}
$\mathcal{M}_1$	$G_1[a]$	...	$G_n[a]$	$G[z^-]$
$\mathcal{M}_2$	$G_1[z]$	...	$G_n[z]$	$G[a]$

Each person in G exists in the same states of nature in $\mathcal{M}_1$ and $\mathcal{M}_2$. $\mathcal{M}_2$ gives each person a better chance of getting at least z, and at least as good a chance of getting any other wellbeing level. Stochastic Dominance for Personal Prospects therefore implies that $\mathcal{M}_2$ is better for each person than $\mathcal{M}_1$. Since both prospects are egalitarian, Weak Pareto for Equal Risk then implies that $\mathcal{M}_2$ is better than $\mathcal{M}_1$.

This is another repugnant conclusion. The G_i sets could each have at least ten billion people, and n could be arbitrarily large. Under these conditions, to say that $\mathcal{M}_2$ is better than $\mathcal{M}_1$ is to say that a near-certainty of having ten billion people with excellent lives (otherwise a very large number of lives barely worth living) is worse than the same near-certainty of there being an enormous number of people with lives barely worth living (otherwise a very large number of excellent lives). This conclusion is hard to square with a rejection of the official Repugnant Conclusion. If we believe that the Repugnant Conclusion is false, we believe that the mere fact that an outcome contains *very many* people is not enough to make that outcome *very good*. Yet in the case of $\mathcal{M}_1$ and $\mathcal{M}_2$, we have to believe that the tiny chance of getting a better outcome in state s_{n+1} is more important than the much larger chance of getting a better outcome in all other states of nature. The only important difference between the comparison between $\mathcal{M}_1$ and $\mathcal{M}_2$ in state s_{n+1} and the comparisons in the other states is that in s_{n+1}, vastly more people exist.[28] The claim that this can make all the difference is precisely the claim we deny when we reject the Repugnant Conclusion.

6 Conclusion

Nebel's Intrapersonal Addition Paradox brings yet more trouble for those who wish to avoid the Repugnant Conclusion. In §3, we saw that provided one accepts the transitivity of the prudential and moral at-least-as-good-as relations, the Repugnant Conclusion can only be robustly avoided by denying the Probable Addition Principle or Weak Pareto for Equal Risk. Because the Probable Addition Principle is the less compelling of the two principles by a significant margin, denying Probable Addition seemed to be the best bet. However, we found in §4 that Probable Addition can be replaced by the much more plausible Conditional Value Principle. The argument demonstrating this required slightly stronger auxiliary assumptions than the Intrapersonal Addition Paradox, but these stronger assumptions are not much less plausible than the originals.

Since denying the Probable Addition Principle is not enough to block the Repugnant Conclusion, the only remaining option is to deny Weak Pareto for Equal Risk. This was confirmed in §5, where we saw that an intuitively "repugnant" conclusion can be derived from Weak Pareto for Equal Risk, together with the compelling principle of Stochastic Dominance for Personal Prospects. Notably, unlike most arguments for versions of the Repugnant Conclusion, this argument did not assume transitivity or any similar principle.

Given that we must choose between avoiding the Repugnant Conclusion and accepting Weak Pareto for Equal Risk, which option is best? I am inclined to favour the second option, mostly because accepting the Repugnant Conclusion provides a response to interpersonal and intrapersonal arguments with a single stroke, whereas if we deny Weak Pareto for Equal Risk, we are still left with the considerable task of responding to interpersonal arguments which do not appeal to Pareto principles.[29]

That said, denying Weak Pareto for Equal Risk remains a serious option. Although it is a deeply compelling principle, it is not satisfied by standard ex post prioritarianism.[30] Perhaps, then, it could be rejected on Prioritarian grounds. Other grounds for the rejection of Weak Pareto for Equal Risk have been proposed by Petra Kosonen, who argues that we should reject this principle if we adopt a decision theory which ignores small probabilities.[31] The rough idea is that small probabilities of many individuals being better off, where these probabilities are below the threshold for discounting and thus may be ignored prudentially, might add up to a large probability above the threshold for discounting that *somebody* will be better off, which cannot be ignored morally.

A final possibility is to deny Weak Pareto for Equal Risk on the basis that it is just a mistake to think that moral and prudential importance must march in lock-step. We might instead think that moral and prudential importance sometimes come apart, either because some things are morally but not prudentially valuable, or because the extent to which lives are morally valuable can depend on the number and nature of the other lives that exist. For instance, we might believe that the moral difference between the existence of ten billion people with excellent lives and the existence of nobody at all is greater than the moral difference between the existence of ten billion excellent lives and the existence of twenty billion lives of the same quality. If so, we sometimes value the existence of excellent lives in a way that is not fully explained by the prudential value of these lives for the people who enjoy them.

In his closing remarks, Nebel suggests something similar:

> We care very strongly about the existence of the things in wonderful lives—things like loving relationships, creative activities, and sophisticated pleasures. But perhaps we do not value these things—primarily, at least—because they are good for the people whose lives contain them. Perhaps we value these things primarily as impersonal goods.[32]

I believe this is the right lesson to draw from the intrapersonal arguments discussed in this paper. It seems to me that if the existence of ten billion excellent lives would be better than the existence of any number of lives barely worth living, then excellent lives must have impersonal as well as personal value. I leave it to the reader to decide whether to apply *modus ponens* or *modus tollens*.

Notes

*. I would like to thank Ralf Bader, Roger Crisp, Todd Karhu, Petra Kosonen, Kacper Kowalczyk and Jakob Lohmar for their valuable comments on this paper.

1. Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), 388.

2. See Charles Blackorby, Walter Bossert, and David Donaldson, "The Axiomatic Approach to Population Ethics," *Politics, Philosophy & Economics* 2, no. 3 (2003): 342–381, Erik Carlson, "Mere Addition and Two Trilemmas of Population Ethics," *Economics and Philosophy* 14, no. 2 (1998): 283–306, Gustaf Arrhenius, "An Impossibility Theorem for Welfarist Axiologies," *Economics and Philosophy* 16, no. 2 (2000): 247–266; Gustaf Arrhenius, "The Very Repugnant Conclusion," in *Logic, Law, Morality: Thirteen Essays in Practical Philosophy in Honour of Lennart Åqvist*, ed. Segerberg Krister and Rysiek Sliwinski (Uppsala: Uppsala University Press, 2003), 167–180; Gustaf Arrhenius, "The Impossibility of a Satisfactory Population Ethics," chap. 1 in *Descriptive and Normative Approaches to Human Behaviour*, ed. Ehtibar N. Dzhafarov and Lacey Perry (Singapore: World Scientific Publishing Co., 2011), 1–26 and Dean Spears and Mark Budolfson, "Repugnant Conclusions," *Social Choice and Welfare* 57, no. 3 (2021): 567–588, among many others.

3. See Jacob M. Nebel, "An Intrapersonal Addition Paradox," *Ethics* 129, no. 2 (2019): 309–343. Intrapersonal arguments in population ethics are also discussed by Teruji Thomas, "Topics in Population Ethics" (DPhil Thesis, University of Oxford, 2016), and by David McCarthy, Kalle Mikkola, and Teruji Thomas, "Utilitarianism With and Without Expected Utility," *Journal of Mathematical Economics* 87 (2020): 77–113.

4. Nebel, "An Intrapersonal Addition Paradox," 314.

5. This definition relies on the as-yet undefined notion of a "neutral life". Those who think that it is possible for a life to be better or worse than non-existence (*Comparativists*) can easily define a neutral life to be one that is equally as good as non-existence for the person in question. But, as we will see in §3, the arguments of this paper will be of more interest to *Non-Comparativists*, who deny that such comparisons are possible. Several definitions of a "neutral life" have been suggested in the literature by authors who do not wish to assume Comparativism; see for example John Broome, "Goodness is Reducible to Betterness: The Evil of Death is the Value of Life," chap. 3 in *The Good and the Economical: Ethical Choices in Economics and Management*, ed. Peter Koslowski (Berlin: Springer, 1993), 78, Gustaf Arrhenius, "Future Generations: A Challenge for Moral Theory" (PhD diss., Uppsala: University Printers, 2000), 20–21, 25, Charles Blackorby, Walter Bossert, and David Donaldson, *Population Issues in Social Choice Theory, Welfare Economics, and Ethics* (New York: Cambridge University Press, 2005), 25 and Ralf M. Bader, "The Asymmetry," chap. 1 in *Ethics and Existence: The Legacy of Derek Parfit*, ed. Jeff McMahan et al. (Oxford: Oxford University Press, 2022), 16. For the purposes of this paper, it should not matter much which definition is adopted.

6. Nebel, "An Intrapersonal Addition Paradox," 315

7. Nebel, 316.

8. Nebel, 317.

9. Like Nebel, I do not assume that wellbeing levels can be represented by real numbers.

10. Following Nebel, I assume the transitivity of the moral and prudential at-least-as-good-as relations throughout this paper. Some philosophers deny (or at least doubt) transitivity due to its role in arguments for the Repugnant Conclusion, notably Larry S. Temkin, "Intransitivity and the Mere Addition Paradox," *Philosophy & Public Affairs* 16, no. 2 (1987): 138–187; Larry S. Temkin, *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning* (Oxford: Oxford University Press, 2012) and Stuart Rachels, "Counterexamples to the Transitivity of *Better Than*," *Australasian Journal of Philosophy* 76, no. 1 (1998): 71–83; Stuart Rachels, "A Set of Solutions to Parfit's Problems," *Noûs* 35, no. 2 (2001): 214–238; Stuart Rachels, "Repugnance or Intransitivity: A Repugnant but Forced Choice," in *The Repugnant Conclusion: Essays on Population Ethics*, ed. Jesper Ryberg and Torbjörn Tännsjö (London: Kluwer Academic, 2004), 163–186.

11. Nebel, "An Intrapersonal Addition Paradox," 319.

12. Nebel, 320.

13. Nebel, 318.

14. Nebel, 322

15. Not everyone takes acceptance of the Repugnant Conclusion to be a last resort. See Yew-kwang Ng, "What Should We Do About Future Generations? Impossibility of Parfit's Theory X," *Economics and Philosophy* 5, no. 2 (1989): 235–253, Michael Huemer, "In Defence of Repugnance," *Mind* 117, no. 468 (2008): 899–933 and Stéphane Zuber et al., "What Should We Agree on about the Repugnant Conclusion?," *Utilitas* 33, no. 4 (2021): 379–383.

16. Many philosophers have found Non-Comparativism independently compelling. See John Broome, *Ethics Out Of Economics* (Cambridge: Cambridge University Press, 1999), 168, Krister Bykvist, "The Benefits of Coming Into Existence," *Philosophical Studies* 135, no. 3 (2007): 335–362, Jeff McMahan, "Causing People to Exist and Saving People's Lives," *The Journal of Ethics* 17, no. 1 (2013): 7 and (seemingly) Parfit, *Reasons and Persons*, 489.

17. This point is not my own. I was made aware of it by Kacper Kowalczyk and Michal Masny, who credit it to an anonymous reviewer of their manuscript, "The Risky Existential Question and the Repugnant Conclusion" (n.d.)

18. Technically, we here need to assume a version of Certainty Equivalence which applies to prudential, rather than overall, value. Otherwise, we only contradict the claim that a *prospect* guaranteeing existence for S cannot be better or worse for S than a prospect guaranteeing the non-existence of S.

19. The canonical formulation of the Sure Thing Principle is given by Leonard J. Savage, *The Foundations of Statistics* (New York: John Wiley & Sons, 1954), 23. My version of the Sure Thing Principle differs from this canonical formulation in several respects. First, it is about prudential value rather than preferences. Second, it pertains to only one state of nature at a time, rather than to arbitrary events (sets of states). Third, it involves conditioning on the non-occurrence of states, rather than uniformly replacing identical pairs of outcomes with other identical pairs of outcomes. The first change is essential, the second change makes my version of the principle weaker, and the third change is unnecessary, but helps to simplify my argument.

20. Consider prospect $\mathcal{P}_1$ from the earlier table. Clearly, $\mathcal{P}_1$ is equally as good as itself. Applying the Sure Thing Principle to ignore State 1, we can conclude that a prospect guaranteeing non-existence is equally as good as itself. This conclusion may be difficult for Non-Comparativists to accept.

21. See Nebel, "An Intrapersonal Addition Paradox," 340–341.

22. This follows from the slightly less minimal principle of Prudence I shall introduce in §4.

23. One might doubt the Conditional Value Principle on the following grounds. Suppose that existence itself has final value, so that additional chances of existence, even at a wellbeing level slightly below neutral, can be better for you. On this view, a conditionally bad prospect with a higher probability of existence might be better for you than a conditionally good prospect which gives you a lower probability of existence. Yet it seems to me that if existence itself is a good, this should be taken into account in the definition of a neutral life, so that a neutral life should be one which contains just enough bad wellbeing components to counterbalance the positive final value of existence. Lives which are bad on this revised definition of neutrality will then contain sufficient bad wellbeing components to outweigh the

final value of existence, so that additional chances of existence at bad wellbeing levels will be worse for the person in question.

24. See Parfit, *Reasons and Persons*, 489.

25. This principle is often called "Priority for Lives Worth Living" in the economics literature. See for instance Blackorby, Bossert, and Donaldson, *Population Issues in Social Choice Theory, Welfare Economics, and Ethics*, 135. I have used a different name in order to avoid giving the false impression that this principle has something to do with Prioritarianism.

26. It seems to me that, provided we countenance comparisons of prospect in risky existence cases at all, there is not much to be said for restricting such comparisons to prospects which involve the same probabilities of existence. (There might be arguments from Non-Comparativism to this conclusion, but if so, I would be inclined to interpret these as arguments against Non-Comparativism.) This leaves open the option of ruling out all comparisons of prospects in risky existence cases, perhaps on the grounds that betterness comparisons are impossible unless the value bearer is guaranteed to exist. It would probably require another paper to give a full assessment of this position; for reasons of space, I shall not discuss it in any detail here. Suffice to say that it does not seem to me particularly plausible, but neither does it seem to be *obviously* wrong. The claims made in this paper should thus be read as being conditional on there being *some* legitimate betterness comparisons in risky existence cases.

27. See Caspar Hare, "Take The Sugar," *Analysis* 70, no. 2 (2010): 237–247. Deference based responses are endorsed by Adam Bales, Daniel Cohen, and Toby Handfield, "Decision Theory for Agents with Incomplete Preferences," *Australasian Journal of Philosophy* 92, no. 3 (2014): 453–470, and by Miriam Schoenfield, "Decision Making in the Face of Parity," *Philosophical Perspectives* 28 (2014): 263–277.

28. Another difference is that the mediocre wellbeing levels in s_1 to s_n are z, rather than the slightly worse level z^- in s_{n+1}. But this difference is not important enough to do the work required. (The mediocre wellbeing levels in s_1 to s_n are only set at z in order for us to be able to say that $\mathcal{M}_2$ is *better* than $\mathcal{M}_1$, rather than the two merely being equally good.)

29. The most concerning interpersonal arguments seem to me those given by Arrhenius, "An Impossibility Theorem for Welfarist Axiologies"; Arrhenius, "The Very Repugnant Conclusion"; Arrhenius, "The Impossibility of a Satisfactory Population Ethics" and by Spears and Budolfson, "Repugnant Conclusions."

30. See Toby Ord, "A New Counterexample to Prioritarianism," *Utilitas* 27, no. 3 (2015): 301.

31. See Petra Kosonen, "Discounting Small Probabilities Solves the Intrapersonal Addition Paradox," *Ethics* 132, no. 1 (2021): 204–217.

32. Nebel, "An Intrapersonal Addition Paradox," 342.