

*Bounded Utilities and Ex Ante Pareto**

Petra Kosonen[†]

May, 2023

ABSTRACT: This paper shows that decision theories on which utilities are bounded, such as standard axiomatizations of Expected Utility Theory, violate Ex Ante Pareto if combined with an additive axiology, such as Total Utilitarianism. A series of impossibility theorems point toward Total Utilitarianism as the right account of axiology, while money-pump arguments put Expected Utility Theory in a favorable light. However, it is not clear how these two views can be reconciled. This question is particularly puzzling if utilities are bounded (as standard axiomatizations of Expected Utility Theory imply) because the total quantity of well-being might be infinite or arbitrarily large. Thus, there must be a non-linear transformation from the total quantity of well-being into utilities used in decision-making. However, such a transformation leads to violations of Ex Ante Pareto. So, the reconciliation of Expected Utility Theory and Total Utilitarianism prescribes prospects that are better for none and worse for some.

*I wish to thank Gustav Alexandrie, Jean Baccelli, Tomi Francis, Andreas Mogensen, Jake Nebel, Teruji Thomas and the audiences of ANU Philosophy Departmental Seminar and the PWI conference ‘Unbounded Utility in Ethics and Decision Theory’ for valuable feedback on this paper.

[†]Population Wellbeing Initiative, University of Texas at Austin. I would be grateful for comments: kosonenpetra@gmail.com

This paper investigates the compatibility of two standard theories: Total Utilitarianism and Expected Utility Theory with a bounded utility function. Let's call the combination of these views *Bounded Expected Totalism*. This paper argues that Bounded Expected Totalism violates *Ex Ante Pareto*, the principle that what is ex ante better for everyone is better overall.^{1,2}

Ex Ante Pareto is often used by utilitarians to justify their theory in opposition to other views, such as prioritarianism and egalitarianism.³ Insofar as Expected Utility Theory is the dominant theory of choice under uncertainty, the argument in this paper could be seen as undermining Total Utilitarianism. So, one might take the argument to support, for example, Average Utilitarianism. On the other hand, the argument may speak differentially in favor of alternatives to Expected Utility Theory, such as expected value maximization.⁴ Another alternative to Expected Utility Theory, discounting small probabilities, also leads to violations of Ex Ante Pareto.⁵ The argument in this paper indirectly supports discounting small probabilities, as it shows that the plausibility of Ex Ante Pareto does not favor Bounded Expected Utility Theory over discounting small probabilities.

The paper proceeds as follows. §1 introduces some background, such as why orthodox decision theory implies bounded utilities. §2 defines Bounded Expected Totalism more formally and explains why it is a *prima facie* attractive view. The

¹There is no inconsistency with Harsanyi's social aggregation theorem. As will be explained later, a bounded expected totalist must reject Harsanyi's conclusion, so they cannot accept all his premises.

²This paper focuses on the compatibility of Expected Utility Theory and Total Utilitarianism, but the problem with Ex Ante Pareto arises for, for example, Critical-Level Utilitarianism in exactly the same way. The problem also arises for Average Utilitarianism and many other theories if individual utilities are unbounded. See for example the argument in Goodsell (2021), which applies to any axiology that is utilitarian in same-number cases. The contribution of this paper is showing that even if utilities are bounded, Total Utilitarianism combined with Expected Utility Theory violates Ex Ante Pareto.

³See Broome (1991, Ch. 9), Parfit (1997), Rabinowicz (2002) and Fleurbaey (2018).

⁴See for example Wilkinson (2022).

⁵Kosonen (2021, §5). On discounting small probabilities and related topics, see Beckstead (2013, Ch 6), Smith (2014), Hájek (2014), Isaacs (2016), Monton (2019), Lundgren & Stefánsson (2020), Beckstead & Thomas (2020), Wilkinson (2022), Russell (2021) and Kosonen (2022).

paper then proceeds to illustrate why this view must violate Ex Ante Pareto. A background issue, which is laid out in §2.2, is how the well-being of a single individual can be traded off between different states of nature. The question is essentially whether the personal value of prospects is risk-averse with respect to well-being. The paper gives separate examples of Ex Ante Pareto violations that involve risk-neutrality (§3) and risk-aversion (§4). §5 concludes by sketching how the examples relate to the classic result in this area, namely, Harsanyi’s social aggregation theorem.

1 Background

This section begins by explaining Total Utilitarianism and Expected Utility Theory. Next, it shows how standard versions of Expected Utility Theory imply bounded utilities. Lastly, it discusses bounded utilities as a possible way of getting intuitively right recommendations in cases that involve tiny probabilities of huge payoffs.

1.1 Total Utilitarianism and Expected Utility Theory

A series of impossibility theorems point toward Total Utilitarianism as the right account of axiology, while money-pump arguments put Expected Utility Theory in a favorable light as a theory of instrumental rationality.⁶ According to Total Utilitarianism, a state of affairs is better than another just in case the total quantity of well-being it contains is greater. The total quantity of well-being in some state of affairs is the sum of individuals’ well-being levels in that state of affairs.

Total Utilitarianism: For all states of affairs A and B , A is at least as good as B if and only if the total quantity of well-being in A is at least as great as the total quantity of well-being in B .⁷

⁶See for example Arrhenius (2000) and Gustafsson (2022b). The impossibility theorems point toward Total Utilitarianism because they show that we cannot escape the Repugnant Conclusion without being forced to accept even more unpalatable conclusions. See also Zuber et al. (2021).

⁷Let $W(A)$ denote the total quantity of well-being in the state of affairs A and let $w(S_i)$ denote

Expected Utility Theory, in turn, states that a prospect is better than another just in case its expected utility is greater.⁸ The expected utility of a prospect is calculated by multiplying the utilities of its outcomes by their probabilities, and summing these up. Let $EU(X)$ denote the expected utility of prospect X . Then, more precisely, Expected Utility Theory states the following:

Expected Utility Theory: For all prospects X and Y , X is at least as good as Y if and only if $EU(X)$ is at least as great as $EU(Y)$.⁹

1.2 Boundedness

Standard versions of expected utility maximization require utilities to be bounded.¹⁰ If utilities are real-valued, then boundedness means the following:

Boundedness: There is some real number M such that for all outcomes x , the utility of x is below M and above $-M$.

In other words, Boundedness rules out arbitrarily and infinitely good outcomes.

Next, consider for example the von Neumann-Morgenstern Expected Utility Theory.¹¹ Let XpY be a risky prospect with a p chance of prospect X obtaining and a $1 - p$ chance of prospect Y obtaining (that is, the agents gets probability p of

the well-being of individual S_i . Then,

$$W(A) = \sum_{i=1}^{\infty} w(S_i).$$

⁸In the case of Total Utilitarianism, ‘better’ is used in an axiological sense; in the case of Expected Utility Theory, ‘better’ is concerned with instrumental rationality.

⁹Let O be the set of possible outcomes, $p_X(o)$ the probability of outcome o in prospect X and $u(o)$ the utility of o . Then,

$$EU(X) = \sum_{o \in O} p_X(o)u(o).$$

¹⁰See for example Kreps (1988, pp. 63–64), Fishburn (1970, pp. 194, 206–207), Hammond (1998, pp. 186–191) and Russell & Isaacs (2021).

¹¹The following axioms together entail Expected Utility Theory: Completeness, Transitivity, Independence and Continuity. See von Neumann & Morgenstern (1947), Jensen (1967, pp. 172–182) and Hammond (1998, pp. 152–164).

prospect X and probability $1 - p$ of prospect Y). Then, if prospects are compared by their expected utilities, Boundedness follows from the following von Neumann-Morgenstern axiom:

Continuity: If X is better than Y , which is better than Z , then there are probabilities p and $q \in (0, 1)$ such that XpZ is better than Y and Y is better than XqZ .

For example, suppose a coin is flipped, and an agent gets X with heads and Z with tails. Suppose further that it is possible to alter the bias of the coin. Continuity requires that, with some bias, the agent prefers the coin flip to certainly getting B , but with some other bias, the agent prefers certainly getting B to the coin flip.

To see why Continuity implies Boundedness (assuming that prospects are compared for their expected utilities), let's consider the two ways in which Boundedness might be false. First, Boundedness might be false because there is an infinite sequence of prospects A_1, A_2, A_3, \dots such that A_2 is at least twice as good as A_1 , A_3 is at least twice as good as A_2 , and so on, with respect to some baseline. For example, suppose A_1 certainly gives two utilities, A_2 four utilities, A_3 eight utilities, and so on. Next, let A be a mixed prospect that assigns probability $1/2^k$ to prospect A_k (that is, probability $1/2$ to prospect A_1 , probability $1/4$ to prospect A_2 , probability $1/8$ to prospect A_3 , and so on). Then, the expected utility of prospect A is infinite:¹²

$$EU(A) = 2 \cdot \frac{1}{2} + 4 \cdot \frac{1}{4} + 8 \cdot \frac{1}{8} + \dots = 1 + 1 + 1 + \dots = \infty.$$

Next, choose some prospects B and C such that B is better than C , and neither is infinitely good or bad. Then, we have that A is better than B (given that A is infinitely good), and B is better than C . However, for all non-zero probabilities p , the expected utility of the mixed prospect ApC is infinite. Therefore, ApC is better than B for all non-zero probabilities p . This is a violation of Continuity; there is

¹²See Peterson (2022) on the St. Petersburg Paradox.

no probability above zero with which ApC is worse than B .¹³

Secondly, and more generally, Boundedness is false if some prospect A is infinitely better than another (good) prospect B . Similarly as above, this leads to a violation of Continuity because the mixed prospect ApC (where C certainly gives nothing) is better than B for all non-zero probabilities p .

To conclude the argument, the two ways in which Boundedness might be false both lead to violations of Continuity. Therefore, it follows from Continuity that Boundedness is true.¹⁴

1.3 Probability Fanaticism

Boundedness has been discussed as a possible alternative to Probability Fanaticism.¹⁵ Probability Fanaticism is the idea that tiny probabilities of large positive or negative payoffs can have enormous positive or negative expected utility (respectively). More formally, it states the following:¹⁶

Probability Fanaticism:

- i *Positive Probability Fanaticism* For any non-zero probability p , and for any (finitely) good outcome o , there is some great enough outcome O such that probability p of O (and otherwise nothing) is better than certainty of o .¹⁷
- ii *Negative Probability Fanaticism* For any non-zero probability p , and for any (finitely) bad outcome $-o$, there is some suffi-

¹³This is a modified argument from Kreps (1988, pp. 63–64).

¹⁴These arguments show that Continuity implies an upper bound on utilities. One can give similar arguments to show that Continuity implies a lower bound on utilities.

¹⁵See for example Beckstead & Thomas (2020).

¹⁶Wilkinson (2022, p. 449). For discussions related to Probability Fanaticism, see Beckstead (2013, ch. 6), Beckstead & Thomas (2020), Goodsell (2021), Russell & Isaacs (2021), Russell (2021) and Wilkinson (2022).

¹⁷In this context, ‘otherwise nothing’ means retaining the status quo or baseline outcome.

ciently bad outcome $-O$ such that probability p of $-O$ (and otherwise nothing) is worse than certainty of $-o$.

How does Boundedness avoid Probability Fanaticism? If utilities are bounded, then sufficiently small probabilities of even very good (or very bad) outcomes do not contribute much to the expected utility of a prospect. For a given probability, there is an upper/lower bound on the contribution to expected utility from outcomes associated with that probability. If the probability gets smaller, this bound also shrinks proportionally so that small enough probabilities cannot help but contribute only a small amount of expected (positive or negative) utility.

For any tiny probability of a great outcome, there is still some certain modest positive outcome that is worse. However, it is not the case that for any certain modest positive outcome, an *arbitrarily* small probability of a sufficiently great outcome is better. If the probability of the great outcome is small enough, increases in the payoff can no longer compensate for decreases in its probability. So, Boundedness prevents such outcomes from dominating the expected utility calculations, and thus, it escapes Probability Fanaticism (assuming fixed upper and lower bounds on utilities).

Let's consider a case that illustrates how Boundedness avoids Probability Fanaticism. Say a case is *fanatical* if tiny probabilities of enormous positive or negative outcomes dominate the expected utility calculations in that case. One example of a fanatical case is *Pascal's Mugging*.¹⁸

Pascal's Mugging: A stranger approaches Pascal and claims to be an Operator from the Seventh Dimension. The stranger promises to perform magic that will help quadrillions of orphans in the Seventh Dimension if Pascal pays the mugger ten livres.

¹⁸Bostrom (2009). The case presented here is a slightly modified version of Bostrom's case. In Bostrom's case, the mugger promises to give Pascal an extra thousand quadrillion happy days of life and help many orphans in the Seventh Dimension. The case is based on informal discussions by various people, including Eliezer Yudkowsky (2007b). Another fanatical case is the St. Petersburg Paradox. See for example Peterson (2022).

Pascal thinks that the mugger is almost certainly lying. However, if utilities are unbounded, the mugger can always increase the payoff until the offer has positive expected utility—at least if Pascal assigns some non-zero probability to the mugger being able and willing to deliver any finite quantity of utility for Pascal.¹⁹ Then, with some number of orphans, the expected-utility-maximizing act is to pay the mugger ten livres. Moreover, the mugger can also ask for more money and increase the payoff accordingly. So, someone who maximizes expected utility with an unbounded utility function would be willing to pay any sum, provided that the payoff is sufficiently large.

In contrast, Bounded Expected Utility Theory has upper and lower bounds on utilities. Consequently, there is an upper limit to how much a bounded expected utility theorist would be willing to pay the mugger (assuming fixed upper and lower bounds on utilities). Bounded Expected Utility Theory does not escape the mugging entirely because, for any payoff offered by the mugger, there is *some* amount a bounded expected utility theorist would pay. After all, a tiny chance of obtaining the upper limit of utilities is worth something. But at least a bounded expected utility theorist would not lose all their money.²⁰ So, Bounded Expected Utility Theory helps avoid the worst instances of Probability Fanaticism (again assuming fixed upper and lower bounds on utilities).

However, this paper shows that Bounded Expected Utility Theory violates Ex Ante Pareto if combined with an additive axiology, such as Total Utilitarianism: The combination of these views prescribes prospects that are better for none and worse for some.

Ex Ante Pareto: For all prospects X and Y , if X is at least as good as Y for everyone, and X is better than Y for some, then X is better than Y .

Bounded Expected Totalism violates Ex Ante Pareto if there is a non-zero prob-

¹⁹Contrary to this, see Hanson (2007), Yudkowsky (2007a) and Baumann (2009).

²⁰This may not be true if the mugger repeatedly returns with the same offer.

ability that an infinite or arbitrarily large number of individuals exist. But it also violates Ex Ante Pareto if it avoids Probability Fanaticism, even if the number of individuals who might exist is capped (as will be explained shortly).

2 Bounded Expected Totalism

This section presents Bounded Expected Totalism in more detail. It also discusses the cardinal structure of well-being, as this issue is relevant to whether individuals can be risk-averse with respect to well-being. Later, §3 and §4 give separate examples of Ex Ante Pareto violations for risk-neutrality and risk-aversion.

2.1 The social transformation function

Let *well-being* refer to how good some outcome is for an individual. And, let *social utility* refer to how good some outcome is overall, from an axiological point of view. Also, let *expected individual utility* represent how good some prospect is for an individual, and let *expected social utility* represent how good some prospect is overall. In the context of Expected Utility Theory, these will be denoted by EU_{Ind} and EU_{soc} , respectively. In general, *individual betterness* will be used to refer to betterness from an individual's point of view. Similarly, *overall/impersonal betterness* will be used to refer to betterness from a moral point of view.

To combine Total Utilitarianism and Expected Utility Theory, we need a *social transformation function* that takes the total quantity of well-being as input and gives social utilities as output. This transformation function must be non-linear if an infinite or arbitrarily large number of happy individuals might exist, as then the total sum of individuals' well-being might be infinite or arbitrarily large (and similarly for negative well-being).²¹ But, as Bounded Expected Totalism requires expected

²¹Note that the total quantity of well-being is not necessarily infinite if an infinite number of individuals exist. For example, suppose that for each individual $k \in \{1, 2, \dots\}$, k 's well-being measure takes a value in the interval $(0, 2^{-k})$. Then, an infinite number of individuals exist but the total quantity of well-being is bounded. However, this can be ruled out by requiring the individual

social utilities to be bounded, the expected social utilities assigned to prospects that might result in an infinite or arbitrarily large number of happy individuals must be bounded.²²

One might object that the total quantity of well-being cannot be infinite or arbitrarily large because there is an upper limit to how many individuals might exist. This upper limit might be due to, for example, the Universe being finite. However, this may not be true, so we need a decision theory that can also handle these possibilities.²³ If there is even a tiny probability that an infinite or arbitrarily large number of individuals exist, then the transformation function must be non-linear for utilities to be bounded. Consider for example the following versions of Pascal's Mugging:

Pascal's Mugging (infinite orphans): The mugger promises to perform magic that will help an *infinite* number of orphans in the Seventh Dimension if Pascal pays the mugger ten livres.

Pascal's Mugging (St. Petersburg case): The mugger promises to perform magic that gives a $1/2$ probability of helping two orphans, a $1/4$ probability of helping four orphans, a $1/8$ probability of helping eight orphans, and so on, if Pascal pays the mugger ten livres.

Suppose Pascal has a non-zero credence in the mugger telling the truth. In that case, he needs to assign some expected social utility to the possibility of helping

well-being measures to have the same range.

²²Beckstead & Thomas (2020, p. 9) write that Boundedness conflicts with the most natural understanding of utilitarianism as an evaluative theory on which improving n lives by a given amount improves the world by n times as much as improving one life. Similarly, they point out that Total Utilitarianism and its variants put unbounded value on creating good lives.

²³As Branwen (2009) put it: "Scientists have suggested infinite universes on multiple occasions, and we cannot rule the idea out on any logical ground. Should our theory of rationality stand or fall on what the cosmologists currently think?" Also, Bostrom (2011, p. 10) writes that recent cosmological evidence suggests that the world is probably infinite, which means that it contains an infinite number of galaxies, stars and planets. And, Bostrom writes, if there are an infinite number of planets, then there is, with probability one, an infinite number of people.

an infinite or arbitrarily large number of orphans. And, if utilities are bounded, then the utility assigned cannot be infinite. Thus, the social transformation function must be non-linear. Moreover, anyone could be confronted with these kind of offers. So, all agents need a theory that can handle cases such as these.

In the previous two cases, the mugger promises to help an infinite number of orphans in expectation, which forces the social transformation function to be non-linear.²⁴ However, even if the mugger does not promise to help an infinite number of individuals in expectation, Bounded Expected Totalism does not avoid Probability Fanaticism if the social transformation function is linear and there is no upper limit to how many individuals might exist. For example, the mugger can always promise to help a greater number of orphans and thus increase the payoff arbitrarily high:

Pascal's Mugging (any number of orphans): The mugger promises to perform magic that will help n number of orphans, where n is finite but arbitrarily large.

If social utilities are linear with the total quantity of well-being, then Bounded Expected Totalism recommends paying the mugger any sum of money, provided that the number of orphans is sufficiently high. That is, for any tiny probability p of the mugger telling the truth, and for any sum of money x , there is some finite number of orphans n , such that Pascal ought to pay the mugger x if the mugger promises to help n orphans. Thus, Bounded Expected Totalism does not avoid Probability Fanaticism if there is no upper limit to how many individuals might exist and the social transformation function is linear.

Lastly, even if we were certain that there is an upper limit to how many individuals might exist, the total quantity of well-being might still be very large. In that case, Bounded Expected Totalism could do with a linear social transformation function, as the requirement for utilities to be bounded would already be satisfied.

²⁴One might object that Total Utilitarianism is not intended to apply in infinite cases. After all, in infinite cases, the total quantity of well-being is not well-defined. So, one might think that Total Utilitarianism does not make sense if there might be an infinite number of individuals.

However, if Bounded Expected Totalism is to avoid fanatical prescriptions in cases that involve tiny probabilities of huge payoffs, then the upper and lower bounds cannot be very high or very low (respectively). So, if a very large number of individuals exist, then the transformation function must be non-linear—or Bounded Expected Totalism does not avoid Probability Fanaticism in an intuitively adequate way.

Bounded Expected Totalism would, technically, avoid Probability Fanaticism if there is an upper limit to how many individuals might exist (and individual utilities are bounded). This is because then it would not be true that, for any certain modest outcome, an arbitrarily small probability of a sufficiently great outcome is better (and similarly for negative outcomes). However, Bounded Expected Totalism would still prescribe what might be considered fanatical choices in cases that involve tiny probabilities of huge outcomes. This happens because the values of those outcomes can be very high (or very low) and, thus, dominate the expected utility calculations. For example, Bounded Expected Totalism might advise Pascal to pay a too high a price to the mugger.

So, there are three reasons to adopt a non-linear social transformation function: First, in expectation, an infinite number of individuals might exist, and these possibilities must be assigned a bounded expected social utility. Secondly, arbitrarily many individuals might exist, in which case Bounded Expected Totalism does not avoid Probability Fanaticism if the social transformation function is linear. Lastly, even if there is an upper limit to how many individuals might exist, the number of possible individuals might be very large. In that case, Bounded Expected Totalism would still prescribe fanatical choices.

Suppose that the social transformation function is non-linear. It will also have the following qualities: First, more well-being is always better, so the social transformation function must be strictly increasing with the total quantity of well-being; it must assign greater utilities to outcomes that contain more well-being. Secondly, because utilities are bounded above, similar increases in well-being must (after some point at least) matter less and less. Consequently, the social transforma-

tion function must be strictly concave on some subset of its domain. Furthermore, because utilities are also bounded below, similar increases in negative well-being must (after some point at least) matter less and less. Thus, the social transformation function must be strictly convex on some subset of its domain. Lastly, for utilities to be bounded, the social transformation function must be sufficiently concave with positive total well-being and sufficiently convex with negative total well-being; the contribution of additional (positive or negative) well-being to social utility must tend to zero.

Let f be this transformation function. Also, let $p(A_i)$ denote the probability of state of affairs A_i and $W(A_i)$ the total quantity of well-being in A_i . Then, we can state Bounded Expected Totalism formally as follows:²⁵

Bounded Expected Totalism: For all prospects X and Y , X is at least as good as Y if and only if $EU_{\text{soc}}(X)$ is at least as great as $EU_{\text{soc}}(Y)$, where

$$EU_{\text{soc}}(X) = \sum_{i=1}^{\infty} p(A_i) f(W(A_i)).$$

Bounded Expected Totalism is the view that outcomes are ranked by their total

²⁵This paper discusses what might be called *Ex-Post Bounded Expected Totalism*. However, there is another way Bounded Expected Totalism can deal with risk, which could be called *Ex-Ante Bounded Expected Totalism*. Formally, *Ex-Ante Bounded Expected Totalism* states the following:

Ex-Ante Bounded Expected Totalism: For all prospects X and Y , X is at least as good as Y if and only if $EU_{\text{soc}}(X)$ is at least as great as $EU_{\text{soc}}(Y)$, where

$$EU_{\text{soc}}(X) = f\left(\sum_{i=1}^{\infty} p(A_i) W(A_i)\right).$$

Ex-Ante Bounded Expected Totalism violates Continuity. For example, let A be a St. Petersburg-style lottery (with the outcomes being total quantities of well-being), B a prospect that certainly gives a modest good outcome and C a prospect that certainly gives nothing. The expected total well-being of the mixed prospect ApC is infinite for all $p \in (0, 1)$. Thus, the expected social utility of ApC equals the upper bound of utilities, which is greater than the expected social utility of B . So, A is better than B , which is better than C , but ApC is better than B for all $p \in (0, 1)$ —which is a violation of Continuity.

quantities of well-being, and prospects are ranked by expected social utilities, where social utility is some bounded function of the total quantity of well-being. On Bounded Expected Totalism, when calculating the value of a prospect, one first calculates the total quantity of well-being in every possible state of the world. Then, one transforms each state's total quantity of well-being into social utilities. Finally, to get the expected social utility of a prospect, one multiplies the social utility of each state with that state's probability and sums these up.

To summarize, social utilities might be bounded if the total quantity of well-being is itself necessarily bounded. However, this is not true; therefore, Bounded Expected Totalism requires a social transformation function that takes the total quantity of well-being as input and outputs social utilities. This social transformation function must be non-linear for three reasons: 1.) An infinite number of individuals exist in expectation, 2.) arbitrarily many individuals might exist or 3.) a very large number of individuals might exist. The latter two reasons apply if one wishes to avoid fanatical prescription in cases that involve tiny probabilities of huge payoffs.

2.2 The cardinal structure of well-being

As mentioned above, the social transformation function takes the total quantity of well-being as input. To make sense of 'total quantity of well-being', we need well-being to have a 'cardinal structure', which allows us to make statements about *how much* more well-being an individual has in some outcome compared to another outcome.²⁶

²⁶Note that in order to talk of 'negative utilities', a cardinal structure is not sufficient; for that, well-being must have a ratio structure—which the von Neumann-Morgenstern axioms cannot deliver. Total Utilitarianism requires a meaningful zero level of well-being, which a merely interval/cardinal scale does not provide.

Well-being as VNM utilities

There are two ways of deriving the cardinal structure of well-being. First, the cardinal structure of well-being might be understood in a ‘primitivist’ sense, according to which it can be defined independently of the individual betterness relation on gambles.²⁷ Alternatively, the cardinal structure of well-being might be understood in a technical sense as, for example, von Neumann-Morgenstern utilities. On the technical understanding, if the individual betterness relation satisfies a set of axioms, it can be represented by an expectational utility function.

Broome suggests that the meaning of our quantitative notion of good (i.e., well-being) must be determined in this way. He proposes that ‘utility’ embodies the results of weighing good across states of nature.²⁸ Broome (1991, p. 147) writes: “To say that two differences in good are the same may mean nothing more than that they count the same when weighed against each other; they are evenly balanced in determining overall good. This would mean that two differences in good are the same whenever the corresponding differences in utility are the same. And that would be enough to ensure that utility is an increasing linear transform of good. Utility, then, would measure good cardinally. [...] In brief, the suggestion is that our metric of good may be determined by weighing across states of nature.”²⁹

If von Neumann-Morgenstern utilities represent the cardinal structure of well-being, then individual betterness is, by definition, risk-neutral with respect to well-being. It might still be risk-averse with respect to money or happy years of life. But it cannot be risk-averse with respect to well-being because well-being just is the quantity whose expectation the betterness relation can be represented as maximiz-

²⁷Greaves (2015).

²⁸Broome (1991, p. 146). Note that we need not equate utility with how much the agent values those gambles (i.e., their preferences). Utilities tell us which gambles are better and worse for a person relative to a given probability assignment, and—especially since the probability assignment at issue need not be the agent’s own—this need not coincide with what the agent prefers.

²⁹Broome (1991, p. 148) also concedes that we might find a metric of well-being in some other way. For example, instead of weighing up across the dimension of states of nature, he writes that this metric might be found by weighing up across a different dimension, such as the dimension of time.

ing. This view satisfies the following principle:³⁰

Bernoulli's hypothesis: One alternative is at least as good for a person as another if and only if it gives the person at least as great an expectation of their well-being.

Bernoulli's hypothesis implies risk-neutrality about well-being.³¹ It also tells us that utility represents well-being cardinally.

This paper focuses mostly on lifetime well-being. But many of the same issues arise when we aggregate intrapersonal well-being over time.³² Let *momentary well-being* mean how good things are for a person at some time. At least in theory, an agent can live infinitely or arbitrarily long at a given level of bliss. Therefore, for well-being/utilities to be bounded, momentary well-being must have diminishing marginal well-being/utility. Additional happy years of life must contribute less the more happy years the agent already has (and similarly for unhappy years of life).

Risk-aversion with respect to well-being

If Bernoulli's hypothesis is false, then individual betterness might be risk-averse with respect to well-being. For example, agents might be represented as maximizing risk-weighted expected utility.³³ Buchak's (2017) representation theorem shows how to construct a cardinal utility function, which may serve as a cardinal well-being measure without committing to the requirement that individual betterness maximizes the expectation of said utility function. Alternatively, well-being could be understood in a primitivist sense. The primitivist view requires that quantities of well-being have meaning independently of how much they count when evaluating uncertain prospects.³⁴ But if such a metric of well-being is available,

³⁰Broome (1991, p. 142). I have replaced 'good' with 'well-being'.

³¹Broome (1991, pp. 124 and 203).

³²See Broome (1991, p. 226) on the *Intertemporal Addition Theorem*, which concerns the aggregation of intrapersonal well-being over time.

³³See for example Quiggin (1982), Buchak (2013) and Buchak (2017).

³⁴Broome (1991, p. 217). For some possible primitivist views, see Greaves (2017, pp. 200–202).

then individual betterness might be risk-averse with respect to this (non-technical) well-being. Note that this view is compatible with Expected Utility Theory (but not with Bernoulli's hypothesis).

Next, let an *agent's transformation function* be a function that takes that person's well-being levels as input and outputs their individual utilities (to be used in decision-making under risk). If individual betterness over prospects is sufficiently risk-averse with respect to well-being, such that the agent's transformation function approaches asymptotically some upper bound with more well-being, then well-being itself can be unbounded without leading to unbounded utilities.

Unbounded well-being with risk-neutrality

Finally, individual betterness might be risk-neutral with respect to well-being. And, happy days of life might not contribute less to well-being the more happy days the agent already has (and similarly for unhappy days). Given that individuals might live arbitrarily long at a constant positive well-being level, this view implies that both well-being and utilities are unbounded. This leads to a prudential analogue of Probability Fanaticism:

Prudential Fanaticism:

- i *Positive Prudential Fanaticism* For any non-zero probability p , and for any (finitely) good outcome o , there is some great enough outcome O such that probability p of O (and otherwise nothing) is prudentially better than the certainty of o for some individual S .
- ii *Negative Prudential Fanaticism* For any non-zero probability p , and for any (finitely) bad outcome $-o$, there is some bad enough outcome $-O$ such that probability p of $-O$ (and otherwise nothing) is prudentially worse than the certainty of $-o$ for some individual S .

To summarize, the social transformation function uses the ‘total quantity of well-being’ as input. To make sense of this notion, well-being must have a cardinal structure. This structure could be primitive, that is, given independently of individual betterness relation on gambles. Alternatively, it could be defined in a technical way, as for example von Neumann-Morgenstern utilities. If the cardinal structure is defined using the von Neumann-Morgenstern axioms, then individual betterness is risk-neutral. But if it is primitive, or defined in some other way, then it is at least initially an open question whether individual betterness is risk-neutral, risk-averse, or what, with respect to well-being. The next section shows that Bounded Expected Totalism violates Ex Ante Pareto if individual betterness is risk-neutral with respect to well-being. §4 shows that the violation happens even if individual betterness is risk-averse with respect to well-being.

3 The risk-neutral case

This section shows that Bounded Expected Totalism violates Ex Ante Pareto if individual betterness is risk-neutral with respect to well-being.

Let well-being levels be represented by real numbers. As argued above, the social transformation function f must be strictly concave on some subset of its domain. For the sake of argument, let’s suppose it is strictly concave at 1. Then, there must be some positive constants δ and ϵ such that $f(1) - f(1 - \delta) > f(1 + \delta + \epsilon) - f(1)$. This is because the smaller benefit (δ) contributes more when added to a population at a lower well-being level than the greater benefit ($\delta + \epsilon$) when added to a population at a higher well-being level.

Next, consider the following prospects:

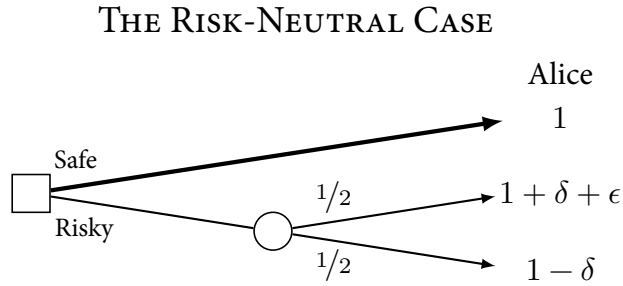
The Risk-Neutral Case:

Risky Gives a 0.5 probability of a $1 + \delta + \epsilon$ well-being level; otherwise, it gives a well-being level of $1 - \delta$.

Safe Surely gives a well-being level of 1.

Suppose that the betterness relation of some agent, Alice, is risk-neutral with respect to her well-being. Then, Risky is better than Safe for Alice (with all positive values of δ and ϵ), as Risky gives her a higher expectation of well-being than Safe does.

But is Risky also better than Safe impersonally? The answer is no. Given that the constants δ and ϵ are such that $f(1) - f(1 - \delta) > f(1 + \delta + \epsilon) - f(1)$, Safe is impersonally better than Risky. The situation is illustrated by the following graph:³⁵



³⁵Gustafsson (2022a) presents this case to illustrate that *Ex-Post* Prioritarianism violates Ex Ante Pareto, a fact that goes back at least to Rabinowicz (2002). For an overview of this topic, see for example Fleurbaey (2018). See also Broome (1991, Ch. 9). Bounded Expected Totalism coincides with *Ex-Post* Prioritarianism in one-person cases. So, we can appeal to the standard fact that *Ex-Post* Prioritarianism violates Ex Ante Pareto. Let $w(S_{ij})$ denote the well-being of individual S_j in A_i . Then, *Ex-Post* Prioritarianism states the following:

***Ex-Post* Prioritarianism:** For all prospects X and Y , X is at least as good as Y if and only if $EU_{\text{Soc}}(X)$ is at least as good as $EU_{\text{Soc}}(Y)$, where

$$EU_{\text{Soc}}(X) = \sum_{i=1}^{\infty} p(A_i) \left(\sum_{j=1}^{\infty} f(w(S_{ij})) \right).$$

Bounded Expected Totalism differs from *Ex-Post* Prioritarianism because it first sums up everyone's well-being and then converts this sum into social utilities. In contrast, the latter view first converts individuals' well-being levels and then sums up the converted well-being levels. Bounded Expected Totalism applies the transformation function to the total quantity of well-being; *Ex-Post* Prioritarianism applies it to the well-being of individuals. On *Ex-Post* Prioritarianism, social utilities are unbounded because the sum of converted well-being levels can be arbitrarily high, given that arbitrarily many individuals might exist. On Bounded Expected Totalism, social utilities are bounded because, although the sum of everyone's well-being can be arbitrarily high, the total quantity of well-being has diminishing marginal utility.

Here, the square represents a choice node, while the circle represents a chance node. Going up at the choice node means accepting Safe, and going down at the choice node means accepting Risky. Thus, if we go up, Alice gets a well-being level of 1. On the other hand, if we go down, there are two possible states of the world, each with a 0.5 probability. In state 1, Alice gets a well-being level of $1 + \delta + \epsilon$. And, in state 2, Alice gets a well-being level of $1 - \delta$.

The expected social utility of going up is $EU_{\text{soc}}(\text{Safe}) = f(1)$. And, the expected social utility of going down is $EU_{\text{soc}}(\text{Risky}) = \frac{1}{2} \cdot f(1 + \delta + \epsilon) + \frac{1}{2} \cdot f(1 - \delta)$. Given that $f(1) - f(1 - \delta) > f(1 + \delta + \epsilon) - f(1)$, $EU_{\text{soc}}(\text{Risky})$ is less than $EU_{\text{soc}}(\text{Safe})$.³⁶ Thus, going up is impersonally better than going down, according to Bounded Expected Totalism. However, going down is better than going up for Alice (and equally good for everybody else). So, we have a violation of Ex Ante Pareto.³⁷

To summarize, Bounded Expected Totalism violates Ex Ante Pareto if individual betterness is risk-neutral with respect to well-being. So, if well-being is understood as von Neumann-Morgenstern utilities, or in a primitive way and individual betterness is risk-neutral with respect to well-being, then the combination of Expected Utility Theory and Total Utilitarianism prescribes prospects that are better

³⁶By rearranging $f(1) - f(1 - \delta) > f(1 + \delta + \epsilon) - f(1)$, we get $f(1) + f(1) > f(1 + \delta + \epsilon) + f(1 - \delta)$. Next, by dividing both sides by 2, we get $f(1) > \frac{1}{2} \cdot f(1 + \delta + \epsilon) + \frac{1}{2} \cdot f(1 - \delta)$.

³⁷If individual utilities are unbounded above while social utilities are bounded above, then Bounded Expected Totalism violates Ex Ante Pareto in the following case as well:

Unbounded individual utilities:

Risky' Gives a tiny probability p of a very high positive well-being level w_1 (and otherwise nothing).

Safe' Surely gives a modest positive well-being level w_2 .

Suppose individuals maximize unbounded expected utility, but social utilities are bounded. Then, with some values of p , w_1 and w_2 , *Risky'* is better than *Safe'* for individuals, but *Safe'* is impersonally better than *Risky'*. This happens because, in the impersonal case, the additional well-being in w_1 is insufficient to compensate for the tiny probability of obtaining it; however, for individual agents, it is sufficient. Bounded Expected Totalism violates Ex Ante Pareto in a similar case (changing what needs to be changed) if individual utilities are unbounded below while social utilities are bounded below.

for none and worse for some. This happens because the social transformation function is concave on some subset of its domain.³⁸ Consequently, overall betterness is at least sometimes risk-averse with respect to (positive) well-being, but individual betterness is always risk-neutral.

4 The risk-averse case

This section shows that Bounded Expected Totalism violates Ex Ante Pareto even if individual betterness is risk-averse with respect to well-being. It is already known that individual risk attitudes incompatible with Expected Utility Theory can cause tensions with Ex Ante Pareto.³⁹ However, the violation of Ex Ante Pareto discussed in this section happens even if the risk-aversion is of the kind that is compatible with Expected Utility Theory.

If individual betterness is risk-averse with respect to well-being, then it may no longer be true that Risky is better than Safe for Alice. So, Bounded Expected Totalism might not violate Ex Ante Pareto in the way discussed earlier. If Alice's transformation function corresponds to the social transformation function when Alice is the only person who exists, then Risky is at least as good as Safe for Alice if and only if Risky is at least as good as Safe impersonally (and vice versa). So, Bounded Expected Totalism avoids violating Ex Ante Pareto in the earlier case.

However, how much Alice's well-being contributes to social utility depends on how many individuals exist and what their well-being levels are. The greater the total quantity of well-being, the smaller the contribution of additional well-being is. Suppose that, when Alice is the only person who exists, Alice's loss of δ would reduce social utility by x units, and her gain of $\delta + \epsilon$ would increase it by more than x units. Then, in the one-person case, Risky is better than Safe (both impersonally

³⁸The same argument can be applied, changing what needs to be changed, as long as the social transformation function is concave on some subset of its domain—it need not be concave specifically at 1.

³⁹See for example Nebel (2020) and Mongin & Pivato (2015).

and, by Ex Ante Pareto, for Alice).⁴⁰

Now change the case; suppose that, besides Alice, there is a large number N of other, unaffected people.

Alice and Others: A large number N of other people have very good lives in state 1 ($p = 0.5$) and neutral lives in state 2 ($p = 0.5$).

Risky Gives Alice a well-being level of $1 + \delta + \epsilon$ in state 1 and a well-being level of $1 - \delta$ in state 2.

Safe Gives Alice a well-being level of 1 in states 1 and 2.

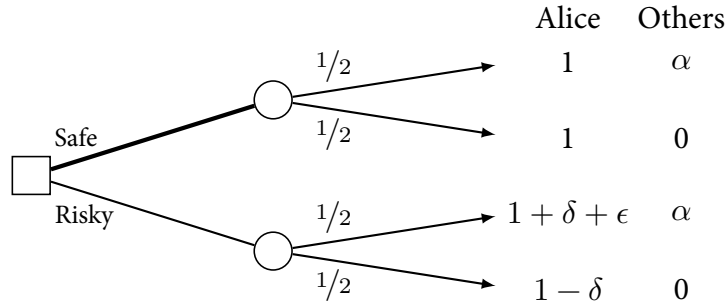
In the state where Alice would lose δ (state 2), the other people have neutral lives (i.e., lives whose addition does not increase or decrease the total quantity of well-being). It follows that, no matter how large N is, her loss of δ would still reduce social utility in that state by x units. On the other hand, in the state where Alice would win $\delta + \epsilon$ (state 1), the N people have very good lives. Let α denote the total quantity of well-being of the N people with very good lives. As we increase N , the social utility in state 1 approaches the upper limit of utilities until it comes within x units of the upper limit. Then, increasing Alice's well-being by $\delta + \epsilon$ contributes less than x to social utility in that state. So, the $\delta + \epsilon$ increase in Alice's well-being in state 1 is no longer sufficient to compensate for the possible loss of δ well-being (and x units of utility) in state 2. It follows that, with a sufficiently large N , Safe is impersonally better than Risky. This contradicts Ex Ante Pareto since Risky is better than Safe for Alice, and Safe and Risky are equally good for each of the N additional people.

⁴⁰Note that this step requires the following version of Ex Ante Pareto:

Weak Ex Ante Pareto: For all prospects X and Y , if X is at least as good as Y for everyone, then X is at least as good as Y .

Also, this step assumes Completeness. Without Completeness, Weak Ex Ante Pareto does not entail that Risky must be better than Safe for Alice if Risky is better than Safe impersonally—they could be incomparable for her.

THE RISK-AVERSE CASE



To summarize, Bounded Expected Totalism violates Ex Ante Pareto even if individual betterness is risk-averse with respect to well-being. So, the combination of Expected Utility Theory and Total Utilitarianism prescribes prospects that are better for none and worse for some. Note that this violation of Ex Ante Pareto happens even if the risk-aversion is of the kind that is compatible with Expected Utility Theory.

5 Harsanyi's social aggregation theorem

This section discusses how the earlier examples relate to a famous result in this area, namely, Harsanyi's social aggregation theorem. Harsanyi's social aggregation theorem shows that if both individual and social betterness relations can be given an expected utility representation, and the overall betterness relation satisfies Ex Ante Pareto, then social utilities are weighted sums of individual utilities.⁴¹

Let me explain Harsanyi's premises in more detail. The first premise states that each individual's betterness relation obeys the von Neumann-Morgenstern axioms.⁴² So, the individual betterness relation can be represented by an expected

⁴¹Harsanyi (1955). Harsanyi (1955) uses individual utilities to describe individual preferences. But we may reinterpret them as describing individual betterness instead of individual preferences. See Broome (1991).

⁴²Harsanyi (1955) uses Marschak's (1950) versions of the von Neumann & Morgenstern (1947) axioms. Marschak's (1950, p. 117) Postulate II states:

tational utility function. The second premise says that the overall betterness relation obeys the von Neumann-Morgenstern axioms. So, overall betterness can also be represented by an expectational utility function. The third premise is Ex Ante Pareto.⁴³ The conclusion of Harsanyi's theorem is that social utilities are weighted sums of individual utilities. Thus, overall betterness can be represented as maximizing the expectation of a weighted sum of individual utilities. If, in addition, we assume equal weighting for all individuals, then this theorem shows that the social utility function must be a sum of individual utilities.⁴⁴

Harsanyi's theorem shows, in other words, that if individual and overall betterness relations are represented by expectational utility functions, then in order to satisfy Ex Ante Pareto, the social utility function must be a linear combination of individual utilities. Earlier in this paper, it was shown that Total Utilitarianism combined with Bounded Expected Utility Theory violates Ex Ante Pareto. Therefore, if one accepts Bounded Expected Totalism, that premise of Harsanyi's theorem fails. The reason that led to its failure was that a non-linear social transformation function is needed because the number of individuals might be infinite or arbitrarily large. In fact, it is unsurprising that one of Harsanyi's premises must be rejected; if the number of individuals might be infinite or arbitrarily large, then social utilities cannot be weighted sums of individual utilities because this could

Postulate II (Continuity): If X is better than Y , which is better than Z , then there is a probability $p \in (0, 1)$ such that Y is equally as good as XpZ .

This postulate implies, in a similar way as shown earlier, that utilities must be bounded.

⁴³Harsanyi (1955) uses Pareto Indifference in the original formulation of the theorem, while Harsanyi (1977, p. 65) uses Weak Ex Ante Pareto:

Pareto Indifference: For all prospects X and Y , if X and Y are equally good for everyone, then X and Y are overall equally good.

Using Weak Ex Ante Pareto instead of Pareto Indifference guarantees that positive individual well-being contributes *non-negatively* to social utilities. Using Ex Ante Pareto instead of Weak Ex Ante Pareto guarantees that positive individual well-being contributes *positively* to social utilities. See Weymark (1994) on Harsanyi's theorem with different Pareto principles.

⁴⁴Broome (1991, §10) argues that Harsanyi's social aggregation theorem, together with Bernoulli's hypothesis, leads to utilitarianism.

lead to unbounded social utilities.^{45,46} So, given that a bounded expected totalist rejects Harsanyi's conclusion, they cannot accept all his premises.

This is worrying because Harsanyi's theorem is often considered one of the best arguments for utilitarianism. The conclusion of Harsanyi's theorem is that, for any fixed and finite population, social utility is an affine (or linear) function of total individual utility. However, once we consider the possibility of an infinite or arbitrarily large population, we find that social utility must be non-linear if social utilities are bounded and additive with individual utilities.⁴⁷ And this leads to violations of Ex Ante Pareto.

All this can be taken to support *Average Utilitarianism*, namely, the view that one population is better than another if and only if the average well-being it contains is greater.⁴⁸ Alternatively, these cases might be taken to undermine Boundedness (and Continuity). One might accept, for example, *Unbounded Expected Totalism*, namely, the view that combines Total Utilitarianism and Expected Utility Theory with an unbounded utility function. However, this view cannot be supported by a version of Harsanyi's theorem that relies on the von Neumann-Morgenstern axiomatization of Expected Utility Theory, as this axiomatization has Continuity as one of its axioms. But one might attempt to justify Unbounded Expected Totalism

⁴⁵See Blackorby et al. (2007) for an extension of Harsanyi's social aggregation theorem to variable populations.

⁴⁶As mentioned earlier, this need not be true. See footnote 21.

⁴⁷Harsanyi (1977, p. 60) himself discusses what he calls the 'boundary problem for the society', namely, whose utility functions ought to be included in our social-welfare function. He considers whether to include, for example, higher animals, distant future generations, robots or the inhabitants of other planets. However, he does not mention the possibility that doing so might lead to infinite or arbitrarily large populations.

⁴⁸Average Utilitarianism does not require a non-linear social transformation function; if individual utilities are bounded, then the average of those must also be bounded. So, Average Utilitarianism avoids violating Ex Ante Pareto. However, Average Utilitarianism has other implausible implications, such as the *Sadistic Conclusion* (Arrhenius 2000, p. 251):

The Sadistic Conclusion: When adding people without affecting the original people's welfare, it can be better to add people with negative well-being rather than positive well-being.

with a Harsanyi-style argument that does not rely on Continuity.⁴⁹

Finally, the arguments in this paper might be taken to indirectly support alternatives to Expected Utility Theory and Boundedness, such as discounting small probabilities. Discounting small probabilities also leads to violations of Ex Ante Pareto.⁵⁰ But given that both Bounded Expected Totalism and discounting small probabilities violate Ex Ante Pareto, the plausibility of Ex Ante Pareto does not favor the former over the latter.

6 Conclusion

This paper has shown that Bounded Expected Totalism violates Ex Ante Pareto. Separate examples of Ex Ante Pareto violations were given for risk-neutrality and risk-aversion. Lastly, the implications of these cases for Harsanyi's social aggregation theorem were discussed. One such implication is that total utilitarians who wish to keep orthodox decision theory can no longer appeal to Harsanyi's argument in support of their theory.

The violations of Ex Ante Pareto happen because there is a non-zero probability that an infinite or arbitrarily large number of individuals exist. But they also happen if one wishes to avoid Probability Fanaticism. Since Bounded Expected Totalism cannot avoid Probability Fanaticism without violating Ex Ante Pareto, these violations of Ex Ante Pareto undermine the plausibility of Bounded Expected Totalism as an alternative to Probability Fanaticism.

To conclude, combining two standard theories, Total Utilitarianism and Expected Utility Theory with a bounded utility function, results in violations of Ex Ante Pareto: The combination of these views implies that a prospect can be impersonally better than another prospect even though it is worse for everyone who is

⁴⁹Fleurbaey (2009) gives such an argument using statewise dominance and anonymity instead of the von Neumann-Morgenstern axioms. Relatedly, McCarthy et al. (2020) show that one can argue for Expected Utility Theory with an unbounded utility function from Pareto and anonymity.

⁵⁰See Kosonen (2021, §5).

affected by the choice.

References

- Arrhenius, G. (2000), 'An impossibility theorem for welfarist axiologies', *Economics and Philosophy* **16**(2), 247–266.
- Baumann, P. (2009), 'Counting on numbers', *Analysis* **69**(3), 446–448.
- Beckstead, N. (2013), On the overwhelming importance of shaping the far future, PhD thesis, Rutgers, the State University of New Jersey.
- Beckstead, N. & Thomas, T. (2020), 'A paradox for tiny probabilities and enormous values'. Global Priorities Institute Working Paper No.10.
URL: <https://globalprioritiesinstitute.org/nick-beckstead-and-teruji-thomas-a-paradox-for-tiny-probabilities-and-enormous-values/>
- Blackorby, C., Bossert, W. & Donaldson, D. (2007), 'Variable-population extensions of social aggregation theorems', *Social Choice and Welfare* **28**(4), 567–589.
- Bostrom, N. (2009), 'Pascal's Mugging', *Analysis* **69**(3), 443–445.
- Bostrom, N. (2011), 'Infinite ethics', *Analysis and Metaphysics* **10**, 9–59.
- Branwen, G. (2009), 'Notes on Pascal's Mugging'.
URL: <https://www.gwern.net/mugging>
- Broome, J. (1991), *Weighing Goods: Equality, Uncertainty and Time*, Blackwell, Oxford.
- Buchak, L. (2013), *Risk and Rationality*, Oxford University Press, Oxford.
- Buchak, L. (2017), 'Precis of *Risk and Rationality*', *Philosophical Studies* **174**(9), 2363–2368.

- Fishburn, P. C. (1970), *Utility Theory for Decision Making*, Wiley, New York.
- Fleurbaey, M. (2009), 'Two variants of Harsanyi's aggregation theorem', *Economics Letters* **105**(3), 300–302.
- Fleurbaey, M. (2018), 'Welfare economics, risk and uncertainty', *Canadian Journal of Economics* **51**(1), 5–40.
- Goodsell, Z. (2021), 'A St Petersburg Paradox for risky welfare aggregation', *Analysis* **81**(3), 420–426.
- Greaves, H. (2015), 'Antiprioritarianism', *Utilitas* **27**(1), 1–42.
- Greaves, H. (2017), 'A reconsideration of the Harsanyi-Sen-Weymark debate on utilitarianism', *Utilitas* **29**(2), 175–213.
- Gustafsson, J. E. (2022a), 'Ex-ante prioritarianism violates sequential ex-ante Pareto', *Utilitas* **34**(2), 167–177.
- Gustafsson, J. E. (2022b), *Money-Pump Arguments*, Cambridge University Press, Cambridge.
- Hájek, A. (2014), 'Unexpected expectations', *Mind* **123**(490), 533–567.
- Hammond, P. J. (1998), Objective expected utility: A consequentialist perspective, in S. Barberà, P. J. Hammond & C. Seidl, eds, 'Handbook of Utility Theory Volume 1: Principles', Kluwer, Dordrecht, pp. 143–211.
- Hanson, R. (2007), 'Pascal's Mugging: Tiny probabilities of vast utilities'.
URL: <https://www.lesswrong.com/posts/a5JAiTdytou3Jg749/pascal-s-mugging-tiny-probabilities-of-vast-utilities?commentId=Q4ACkdYFETHA6EE9P>
- Harsanyi, J. C. (1955), 'Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility', *Journal of Political Economy* **63**(4), 309–321.

- Harsanyi, J. C. (1977), *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge University Press, Cambridge.
- Isaacs, Y. (2016), 'Probabilities cannot be rationally neglected', *Mind* **125**(499), 759–762.
- Jensen, N. E. (1967), 'An introduction to Bernoullian utility theory: I. Utility functions', *The Swedish Journal of Economics* **69**(3), 163–183.
- Kosonen, P. (2021), 'Discounting small probabilities solves the Intrapersonal Addition Paradox', *Ethics* **132**(1), 204–217.
- Kosonen, P. (2022), *Tiny Probabilities of Vast Value*, PhD thesis, University of Oxford.
- Kreps, D. M. (1988), *Notes on the Theory of Choice*, Westview Press, Boulder.
- Lundgren, B. & Stefánsson, H. O. (2020), 'Against the *De Minimis* principle', *Risk Analysis* **40**(5), 908–914.
- Marschak, J. (1950), 'Rational behavior, uncertain prospects, and measurable utility', *Econometrica* **18**(2), 111–141.
- McCarthy, D., Mikkola, K. & Thomas, T. (2020), 'Utilitarianism with and without expected utility', *Journal of Mathematical Economics* **87**, 77–113.
- Mongin, P. & Pivato, M. (2015), 'Ranking multidimensional alternatives and uncertain prospects', *Journal of Economic Theory* **157**, 146–171.
- Monton, B. (2019), 'How to avoid maximizing expected utility', *Philosophers' Imprint* **19**(18), 1–24.
- Nebel, J. M. (2020), 'Rank-weighted utilitarianism and the veil of ignorance', *Ethics* **131**(1), 87–106.
- Parfit, D. (1997), 'Equality and priority', *Ratio* **10**(3), 202–221.

- Peterson, M. (2022), The St. Petersburg Paradox, in E. N. Zalta, ed., ‘The Stanford Encyclopedia of Philosophy’, Summer 2022 edn, Metaphysics Research Lab, Stanford University.
- Quiggin, J. (1982), ‘A theory of anticipated utility’, *Journal of Economic Behavior and Organization* **3**(4), 323–343.
- Rabinowicz, W. (2002), ‘Prioritarianism for prospects’, *Utilitas* **14**(1), 2–21.
- Russell, J. S. (2021), ‘On two arguments for fanaticism’. Global Priorities Institute Working Paper 17–2021.
URL: <https://globalprioritiesinstitute.org/on-two-arguments-for-fanaticism-jeff-russell-university-of-southern-california/>
- Russell, J. S. & Isaacs, Y. (2021), ‘Infinite prospects’, *Philosophy and Phenomenological Research* **103**(1), 178–198.
- Smith, N. J. J. (2014), ‘Is evaluative compositionality a requirement of rationality?’, *Mind* **123**(490), 457–502.
- von Neumann, J. & Morgenstern, O. (1947), *Theory of Games and Economic Behavior*, 2 edn, Princeton University Press, Princeton.
- Weymark, J. A. (1994), Harsanyi’s social aggregation theorem with alternative Pareto principles, in W. Eichhorn, ed., ‘Models and Measurement of Welfare and Inequality’, Springer, Berlin, Heidelberg, pp. 869–887.
- Wilkinson, H. (2022), ‘In defence of fanaticism’, *Ethics* **132**(2), 445–477.
- Yudkowsky, E. (2007a), ‘A comment on Pascal’s Mugging: Tiny probabilities of vast utilities’.
URL: <https://www.lesswrong.com/posts/a5JAiTdytou3Jg749/pascal-s-mugging-tiny-probabilities-of-vast-utilities?commentId=kqAKXskjohx4SSyp4>

Yudkowsky, E. (2007*b*), ‘Pascal’s Mugging: Tiny probabilities of vast utilities’.

URL: <http://www.overcomingbias.com/2007/10/pascals-mugging.html>

Zuber, S., Venkatesh, N., Tännsjö, T., Tarsney, C., Stefánsson, H. O., Steele, K., Spears, D., Sebo, J., Pivato, M., Ord, T., Ng, Y.-K., Masny, M., MacAskill, W., Kuruc, K., Hutchinson, M., Gustafsson, J. E., Greaves, H., Forsberg, L., Fleurbaey, M., Coffey, D., Cato, S., Castro, C., Campbell, T., Budolfson, M., Broome, J., Berger, A., Beckstead, N. & Asheim, G. B. (2021), ‘What should we agree on about the repugnant conclusion?’, *Utilitas* **33**(4), 379–383.