Introduction

Tiny Probabilities of Vast Value

ABSTRACT: This chapter explores different approaches to cases that involve tiny probabilities of huge payoffs. The main approaches discussed are Probability Fanaticism, Boundedness and Probability Discounting. First, the chapter discusses two arguments for maximizing expected utility: the long-run argument and representation theorems. Next, it investigates Probability Fanaticism, on which tiny probabilities of huge positive or negative payoffs can have enormous positive or negative expected utility (respectively). Various arguments for and against Probability Fanaticism are discussed. Then, the chapter considers Boundedness, namely, the idea that utilities are bounded. Finally, the chapter discusses Probability Discounting, on which tiny probabilities should be ignored in practical decision-making. Some other approaches are also discussed briefly. The chapter concludes that the paradoxes involving tiny probabilities of vast value show that some intuitively compelling principles of rationality must be given up.

1 Pascal's Hell

In the beginning, on a small planet in the Solar System, in the Milky Way galaxy…

Satan: I have an offer for you, Pascal, as I have heard that you might be interested in a small probability of a huge payoff.

Pascal: Anything that maximizes expected utility!

Satan: Great! And your utility function is unbounded, am I right?[1]

Pascal: Yes, and additive in terms of people's happy days of life.

Satan: Excellent. So, the offer is this: I will flip a coin, and if it lands on heads, I will help humanity settle on new planets in faraway galaxies and live in bliss until the heat death of the Universe.[2] Until the heat death happens, it will be like heaven. But if the coin lands on tails, then everyone on Earth will suffer excruciating pain for the next fifty years. That's the offer. If you decide to accept it, I will return to Earth every fifty years and give the same offer until either you (or your descendants) refuse the offer, the coin lands on heads, or the Sun expands and makes life on Earth impossible. If you decide not to accept the offer, humanity will live its earthly existence as mere mortals until life on this planet is no longer possible (humanity will not be able to expand out from Earth without my help!)

Pascal: Your offer sounds great—even odds of Utopia! And if we don't win this time, we'll almost certainly win eventually.

Satan: Oh, pardon me, I forgot to say that my coin is somewhat biased. If you

[1] The utility function does not necessarily have to be unbounded for this case to work—it is enough that the upper bound is very high and the lower bound very low. Chapter 1 of this thesis shows that standard axiomatization of Expected Utility Theory require a bounded utility function.

[2] This is the fate of the Universe in which Pascal and Satan live.

accept all the (20 million) offers, the probability of heads happening at least once is one-in-a-googolplex. I admit the odds aren't great. But if the coin lands on heads, I will create a thousand googolplex happy Earth-like planets.

Pascal: Not to worry, the offer is still amazing. The expected value of taking those gambles is clearly greater than the expected value of rejecting them. Actually, its expected value might even be greater than the expected value of the offer I initially thought you were making… So, I'm positively surprised.

Satan: Oops, I made a mistake. I read the wrong page. The instruction manual (*Creating Hell*) says that the probability of heads ever happening on Earth is one-in-Graham's-number. But it is in my power to create any finite number of happy Earth-like planets, so I believe I can still give you a good offer. If the coin lands on heads, I will create a million Graham's number of happy Earth-like planets.[3]

Pascal: Now your offer is even better! Although I dread the almost certain torture for everyone on Earth for the next billion years, the expected value of your offer is far greater than the expected value of not taking it. So, rationality compels me to accept it.

Pascal and Satan then agree on the deal, and Satan flips the coin. Unsurprisingly, it lands on tails.

Satan: You and everyone on Earth will now suffer excruciating pain for the next fifty years.

Pascal: Oh well. I made the right choice, given the information I had. And the future is still great in expectation. Thank you for your offer.

[3] The Universe Pascal and Satan live in is much larger than our Universe.

Satan: I'm always happy to help. See you again in fifty years!

Pascal: See you in fifty (long) years! You are always welcome here.

Satan: I never imagined persuading people to enter (finite) hell would be this easy…

* * *

So Satan traveled from one planet to another, and the inhabitants of those planets—also expected utility maximizers with unbounded utilities—always accepted his offer. And they all lived happily ever after (in expectation). But according to Satan's instruction manual, the probability of the coin ever landing on heads was merely one-in-a-googolplex, so the Universe was almost certain to be void of joy and laughter.[4]

2 Maximizing expected utility

The topic of this thesis is how we should treat tiny probabilities of vast value. This chapter goes over different possible approaches. I will start by considering the idea that rational agents maximize expected utility. Two arguments for maximizing expected utility will be discussed: the long-run argument and representation theorems. I will then present two puzzling cases that involve tiny probabilities of huge

[4]This dialogue is based on Pascal's Mugging by Bostrom (2009), which in turn is based on informal discussions by various people, including Yudkowsky (2007*b*). Pascal's Mugging is similar to Pascal's Wager, except that the former does not involve infinite utilities. Pascal (1958) famously argued that one should believe in God because of the possibility of gaining an infinitely good payoff in Heaven: "Let us weigh the gain and the loss in wagering that God is. Let us estimate these two chances. If you gain, you gain all; if you lose, you lose nothing. Wager, then, without hesitation that He is."

payoffs. The subsequent sections explore expected-utility maximization with unbounded and bounded utility functions, as well as alternatives to expected-utility maximization.

2.1 The long-run argument for maximizing expected utility

According to standard decision theory, a rational agent always maximizes expected utility. An act's expected utility is calculated by summing the utilities of its possible outcomes weighted by their probabilities of occurring, where 'utility' measures how preferable (or valuable) some outcome is compared to the alternatives. Let EU (X) denote the expected utility of prospect X, and let $X \succsim Y$ mean that X is at least as good as Y. Also, let 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, more formally, Expected Utility Theory states the following:

> **Expected Utility Theory:** For all prospects X and Y, $X \succsim Y$ if and only if EU(X) $\geq$ EU(Y), where
>
> $$\mathrm{EU}(X) = \sum_{o \in O} p_X(o) u(o).$$

Why should one accept Expected Utility Theory? One argument for maximizing expected utility—the long-run argument—states that expected-utility maximization is the best policy in the long run. This is because, in the long run, the average amount of utility gained per trial is overwhelmingly likely to be close to the expected value of an individual trial.[5] However, it is not certain that the average utility gain per trial would be close to the game's expected utility—it is merely highly likely.[6] Thus, the argument must be that expected-utility maximization is overwhelmingly likely to be the overall best policy. And, if something is overwhelmingly likely to be the overall best policy, then one should do that. So, one should maximize expected utility.

The long-run argument only works under certain further assumptions about what sorts of gambles will arise in the long run. For example, in Pascal's Hell, *not* maximizing expected utility is overwhelmingly likely to be the overall best policy. So, by the same argument, one should *not* maximize expected utility in this case. Thus, the principle from which the long-run argument gets its intuitive support recommends against expected-utility maximization in some cases. And, the true principles of rationality (if there are any) should apply even in hypothetical cases such as Pascal's Hell. Expected-utility maximization might be overwhelmingly likely to be the overall best policy for us. But it is not so always and for everyone. If one accepts the principle that one should choose whatever policy is overwhelmingly likely to be best overall (either for oneself or the group of all agents), then, under some circumstances, one should not maximize expected utility. So, some other argument is needed to establish that one should always do so.[7]

[5] Briggs (2019).

[6] The *Strong Law of Large Numbers* implies that, for any arbitrarily small real umber $\epsilon > 0$, the probability that the average payoff of a prospect falls within ϵ of its expected utility converges to 1 as the number of trials increases. In other words, as the sample size goes to infinity, the average gain per trial will become arbitrarily close to the prospect's expected utility with probability 1. So, in the long run, the average utility associated with a prospect is virtually certain to equal its expected utility. See Briggs (2019, §2.1).

[7] See Briggs (2019, §2.1) for more discussion of the long-run argument for expected-utility

2.2 Representation theorems

Another argument for maximizing expected utility relies on representation theorems, such as the von Neumann-Morgenstern axiomatization of Expected Utility Theory. This representation theorem shows that the following axioms together entail Expected Utility Theory: Completeness, Transitivity, Independence and Continuity.[8] Let $X \succ Y$ mean that X is strictly preferred (or simply 'preferred') to Y.[9] Also, let XpY be a risky prospect with a p chance of prospect X obtaining and a $1 - p$ chance of prospect Y obtaining. These axioms then state the following:

> **Completeness:** $X \succsim Y$ or $Y \succsim X$.
>
> **Transitivity:** If $X \succsim Y \succsim Z$, then $X \succsim Z$.
>
> **Independence:** If $X \succ Y$, then $XpZ \succ YpZ$ for all probabilities $p \in (0, 1]$.
>
> **Continuity:** If $X \succ Y \succ Z$, then there are probabilities p and $q \in (0, 1)$ such that $XpZ \succ Y \succ XqZ$.

Agents who conform to said axioms can be represented as maximizing expected utility. The argument for expected-utility maximization from representation theorems states that these axioms are the axioms of rational preference.[10] Thus, rational agents can be represented as maximizing expected utility. But why

maximization.

[8] See von Neumann and Morgenstern (1947), Jensen (1967, pp. 172–182) and Hammond (1998, pp. 152–164).

[9] Some prospect X is strictly preferred to another prospect Y when X is weakly preferred to Y, but Y is not weakly preferred to X.

[10] Briggs (2019) and Zynda (2000).

should one think these are the axioms of rational preference? First, one might consider them intuitively plausible, so it might seem intuitively right that rational agents satisfy these axioms. Alternatively, they can be supported by money-pump arguments. A money-pump argument intends to show that agents who violate some alleged requirement of rationality are vulnerable to making a combination of choices that leads to a sure loss. There are money-pump arguments for Completeness, Transitivity, Independence and Continuity.[11] So, if vulnerability to this sort of exploitation is a sign of irrationality, then one ought to satisfy the axioms that together entail Expected Utility Theory. Thus, rational agents maximize expected utility.[12]

2.3 Tiny probabilities of vast utilities

However, maximizing expected utility seems to lead to counterintuitive choices in cases that involve tiny probabilities of huge payoffs (at least if utilities are unbounded or if the upper bound is very high or the lower bound is very low). One such case was presented earlier. It is based on the following case:[13]

> **Pascal's Mugging:** A stranger approaches you and promises to use magic that will give you a thousand quadrillion happy days in the Seventh Dimension if you pay him a small amount of money.

Should you pay the stranger? There is a very small but non-zero probability that

[11] See Gustafsson (forthcoming).

[12] As discussed later, these axioms imply a bounded utility function.

[13] Bostrom (2009). This case is based on informal discussions by various people, including Eliezer Yudkowsky (2007*b*).

the stranger is telling the truth. And if he is telling the truth, then the payoff is enormous. Provided the payoff is sufficiently great, the expected utility of paying the stranger is greater than that of keeping the money. Also, if you have a non-zero credence in him being able and willing to deliver any finite amount of utility, then he can always increase the payoff until the offer has positive expected utility, at least if your utilities are unbounded.[14] So, someone who maximizes expected utility with an unbounded utility function (or with a high enough upper bound) would pay the stranger—which seems counterintuitive.

Another case that involves tiny probabilities of huge payoffs is the St. Petersburg paradox introduced by Nicolaus Bernoulli:[15]

> **St. Petersburg Game:** A fair coin is flipped until it lands on heads.[16]
>
> The prize is then $\$2^n$, where n is the number of coin flips.

While Pascal's Mugging involves large finite payoffs and a finite number of possi-

[14]Contrary to this, Baumann (2009, p. 447) argues that the larger the payoff the mugger promises to deliver, the lower the probability you should assign to the proposition that he will stick with his promise. Moreover, Baumann (2009, p. 447) argues that your probabilities should go down faster than the stranger's offer's utilities go up. Relatedly, Robin Hanson has suggested that in a scenario in which many individuals exist, they cannot all have total control over each other's existence. So, your credence in being able to influence them all should be penalized in proportion to the number of individuals that exist. Thus, credences in the mugger telling the truth should decrease in proportion to the possible payoff. See Hanson (2007) and Yudkowsky (2007*a*). But, in the version of Pascal's Mugging presented in this chapter, the stranger promises to prolong *your* life rather than also help very many orphans (as in Bostrom's version). And, it is less surprising to be in a special position to have so much control over one's future self.

[15]Nicolaus Bernoulli originally proposed a version of this game in 1713. The game was simplified by Gabriel Cramer in 1728 and published by Daniel Bernoulli in 1738. See Pulskamp (2013) and Bernoulli (1954). There are variants of the St. Petersburg game that do not seem to make any sense by the lights of Expected Utility Theory because they have no unique expected utility. See, for example, Nover and Hájek (2004) on the Pasadena game.

[16]What happens if the coin never lands on heads? We may suppose that, in that case, the player wins nothing. As this is a zero-probability event, it does not affect the expected utility of the game. See Chapter 2 of this thesis on Expected Utility Theory and possible states of zero probability.

ble outcomes, the St. Petersburg game involves arbitrarily large finite payoffs and infinitely many possible outcomes. The St. Petersburg game has infinite expected monetary value, so an agent who maximizes expected monetary value would pay any finite amount to play it. But again, this seems counterintuitive. As Nicolaus Bernoulli (agreeing with his friend Gabriel Cramer) writes: "[T]here is no person of good sense who wished to give merely 20 coins."[17] Daniel Bernoulli (cousin of Nicolaus Bernoulli) argues that the expected utility of the game is finite because of the diminishing marginal utility of money.[18] However, one can change the game slightly to bypass this objection by changing the prize from money to something with no diminishing marginal utility, such as (possibly) days of life.[19]

To summarize, Expected Utility Theory states that rational agents maximize expected utility. Expected Utility Theory can be supported with the long-run argument, on which one should maximize expected utility because it is overwhelmingly likely to be the best policy in the long run. Alternatively, Expected Utility Theory can be supported by representation theorems. This argument states that the axioms of Expected Utility Theory are the axioms of rational preference. However, maximizing expected utility seems to lead to counterintuitive choices in cases that involve tiny probabilities of huge payoffs, such as Pascal's Hell, Pascal's Mugging and the St. Petersburg paradox. Expected-utility maximization gives counterin-

[17] Pulskamp (2013, p. 6).

[18] Bernoulli (1954).

[19] Monton (2019, p. 2). This is related to the *Super St-Petersburg Paradox* which Samuelson (1977, p. 32) attributes to Menger (1934) (see Menger [1967] for an English translation). Menger (1967, pp. 217–218) shows that if utilities are unbounded, one can always create a Super St-Petersburg game, in which the payoffs grow sufficiently fast so that the expected utility of the game is infinite.

tuitive recommendations in such cases if utilities are unbounded or if the upper bound is very high or the lower bound very low. The next section discusses an implication of expected-utility maximization with an unbounded utility function; the subsequent section explores expected-utility maximization with a bounded utility function. The idea that tiny probabilities should be ignored in practical decision-making is investigated in §5. Finally, §6 briefly discusses some other approaches.

3 Probability Fanaticism

This section discusses arguments for and against Probability Fanaticism, namely, the idea that we should let tiny probabilities of vast utilities dominate the expected utility calculations. As we will see, there are strong arguments for and against Probability Fanaticism, as some plausible principles support this idea while others undermine it.

3.1 The Continuum Argument for Probability Fanaticism

There seems to be something wrong with a theory that lets tiny probabilities of huge payoffs dictate one's course of action. It might even seem *fanatical*. Thus, we may call this view *Probability Fanaticism*. Probability Fanaticism is the idea that tiny probabilities of huge positive or negative payoffs can have enormous positive or negative expected utility (respectively). Formally, it states the following:[20]

Probability Fanaticism:

[20] Wilkinson (2022, p. 449). Beckstead and Thomas (2020) call this 'Recklessness'.

i *Positive Probability Fanaticism* For any probability $p > 0$, and for any finite utility u, there is some large enough utility U such that probability p of U (and otherwise nothing) is better than certainty of u.[21]

ii *Negative Probability Fanaticism* For any probability $p > 0$, and for any finite negative utility $-u$, there is some large enough negative utility $-U$ such that probability p of $-U$ (and otherwise nothing) is worse than certainty of $-u$.

Probability Fanaticism is supported by a *Continuum Argument*.[22] Consider for example the following case:[23]

> **Devil at Your Deathbed:** You have one year of life left. But the devil appears and offers you ten years of happy life instead, with probability 0.999. You accept the offer. But the devil then offers you 100 years of happy life instead, with probability 0.999^2—just 0.1% lower. After some 50,000 trades, you find yourself with a $0.999^{50,000}$ probability of $10^{50,000}$ years of happy life. Predictably, you die shortly thereafter.

In this case, each deal seems better than the one before. Accepting each deal massively increases the payoff while decreasing its probability by a tiny percentage. However, accepting all trades means trading a certain good payoff (one year of happy life) for an extremely tiny probability of a great payoff.

[21] In this context, 'otherwise nothing' means retaining the status quo or baseline outcome.
[22] This argument is from Beckstead (2013, §6) and Beckstead and Thomas (2020, §1).
[23] Beckstead and Thomas (2020, pp. 4–5)

If p is a probability and n is a number of happy lives, then let $p \cdot n$ be a prospect that gives probability p of n happy lives (and otherwise nothing). Then, the following principle supports accepting all the trades:[24]

> **Anti-Timidity:** For any probabilities $p \gg q$ and numbers of happy lives $N \gg n$, $p \cdot (n + N) \succ (p + q) \cdot n$.

Anti-Timidity says that one can always compensate for a tiny decrease in the probability of a good outcome by increasing the payoff sufficiently. Anti-Timidity is plausible. However, it implies (Positive) Probability Fanaticism; repeated applications of Anti-Timidity (together with transitivity) tell us that a tiny probability of a great payoff is better than certainty of a good payoff.[25] Whichever payoff one starts with, and for any tiny probability $p > 0$, there is some great enough payoff such that probability p of the great payoff (and otherwise nothing) is better than certainty of the original payoff. So, to deny Probability Fanaticism, one must reject Anti-Timidity or transitivity—yet both seem intuitively compelling.

[24] Russell (2021, p. 7) and Beckstead and Thomas (2020, p. 2).

[25] A similar argument can be given to support Negative Probability Fanaticism. Instead of Anti-Timidity, this argument uses the following principle:

> **Negative Anti-Timidity:** For any probabilities $p \gg q$ and numbers of unhappy lives $N \gg n$, $(p + q) \cdot n \succ p \cdot (n + N)$.

3.2 More is Better and Simple Separability imply Probability Fanaticism

Another argument for Probability Fanaticism is that it follows from two plausible principles, namely, More is Better and Simple Separability.[26] More is Better states the following:[27]

> **More is Better:** For probabilities $p \gg q$ and numbers $N \gg n$, $p \cdot N \succ q \cdot n$.

More is Better states that it is better to have a much higher probability of many more happy lives than a smaller probability of fewer happy lives.

Let X be a prospect that concerns what is going on in the part of the world we might make any difference to, and let Y be a prospect that concerns what happens somewhere far away, such as a distant galaxy. Also, let $X \oplus Y$ be the combined prospect of the near prospect X and the far prospect Y. Finally, let a 'simple prospect' be a prospect that has only a finite number of possible outcomes. Then, Simple Separability states the following:[28]

> **Simple Separability:** For all simple near prospects X and Y, and any simple far prospect Z, $X \succ Y$ if and only if $X \oplus Z \succ Y \oplus Z$.

Denying Simple Separability means that uncertainty over what happens in distant places can be relevant to what we ought to do, even when we cannot affect what

[26] This argument is also from Beckstead and Thomas (2020, §3.2). The presentation follows closely Russell (2021, §2).

[27] Russell (2021, p. 6).

[28] Russell (2021, p. 15).

happens in those distant places.

To see how More is Better and Simple Separability imply Probability Fanaticism, consider the following prospects:

> **More vs. Less:** Let $p \gg q$ and $N \gg n$. Also, let the probabilities of states 1, 2 and 3 be p, q and $1 - p - q$ (respectively).
>
> *More* Gives N happy lives in state 1 and nothing in states 2 and 3.
>
> *Less* Gives n happy lives in state 2 and nothing in states 1 and 3.

Suppose you face a choice between More and Less, while the inhabitants of a distant Earth-like planet face the following prospect:

> *Far* Gives n happy lives in state 1 and nothing in states 2 and 3.

Given that the Earth-like planet faces prospect Far, the choice you face is between More $\oplus$ Far and Less $\oplus$ Far (see table 1). And, given that More is better than Less (by More is Better), it follows by Simple Separability that More $\oplus$ Far is better than Less $\oplus$ Far. However, as seen in table 1, More $\oplus$ Far gives a slightly lower probability p of a much large number of happy people $n + N$. Thus, More is Better and Simple Separability imply Anti-Timidity: A slightly smaller probability of a much large number of happy lives is better than a slightly higher probability of many fewer happy lives. And, as we saw in the previous section, Anti-Timidity (together with transitivity) implies Probability Fanaticism. Therefore, More is Better and Simple Separability (together with transitivity) imply Probability Fanaticism. To deny Probability Fanaticism, one must reject More is Better, Simple Separability or transitivity—yet they all seem intuitively compelling.

Table 1

More ⊕ Far vs. Less ⊕ Far

	State 1	**State 2**	**State 3**
Probability	p	q	$1-p-q$
More ⊕ Far	$n+N$	0	0
Less ⊕ Far	n	n	0

3.3 Stochastic Dominance and Simple Separability imply Probability Fanaticism

Another related argument for Probability Fanaticism is that it follows from Simple Separability and another very compelling principle, namely, Stochastic Dominance.[29] Let $X = \{x_1,\, p_1;\; x_2,\, p_2\,; \dots\}$ stand for prospect X that gives non-zero probabilities p_1, p_2, and so on, of outcomes x_1, x_2, and so on. Stochastic Dominance then states the following:[30]

Stochastic Dominance: For all prospects $X = \{x_1,\, p_1;\; x_2,\, p_2\,; \dots\}$ and $Y = \{y_1,\, q_1;\; y_2,\, q_2\,; \dots\}$, X is at least as good as Y if, for all out-

[29]The presentation follows closely Russell (2021, pp. 30–33). See Wilkinson (2022, §VI A) for a very similar argument. Also see Tarsney (2020), Beckstead and Thomas (2020) and Goodsell (2021).

[30]Buchak (2013, p. 42). More precisely, this is *first-order stochastic dominance*, an idea that was introduced to statistics by Mann and Whitney (1947) and Lehmann (1955), and to economics by Quirk and Saposnik (1962). The name 'first-degree stochastic dominance' is due to Hadar and Russell (1969, p. 27).

comes o,

$$\sum_{\{i \mid x_i \succsim o\}} p_i \geq \sum_{\{j \mid y_j \succsim o\}} q_j.$$

If in addition, for some outcome u,

$$\sum_{\{i \mid x_i \succsim u\}} p_i > \sum_{\{j \mid y_j \succsim u\}} q_j,$$

then X is better than Y.

One violates Stochastic Dominance if, for all outcomes, some prospect X gives an at least as high probability of an at least as great outcome as some other prospect Y does, but X is not judged at least as good as Y. One also violates Stochastic Dominance if, in addition, X gives a greater probability of an at least as great outcome as Y does for some outcome—yet X is not judged better than Y.

To see how Simple Separability and Stochastic Dominance imply Probability Fanaticism, consider the following prospects:

Safe vs. Risky:

Safe Certainly gives one happy life.

Risky Gives probability $p > 0$ of $n + 1$ happy lives (a great outcome) and otherwise nothing.

Suppose p is tiny. Then, the comparison between Risky and Safe can be considered at a more abstract level whereby it simply corresponds to Positive Probability

Fanaticism. So, Probability Fanaticism is true if Risky is better than Safe. Also, suppose you face the choice between Safe and Risky, while the inhabitants of a distant Earth-like planet face the following prospect (see table 2):

Twin Earth Gives p chance of nothing, q chance of one happy life, q chance of two happy lives, q chance of three happy lives, … , q chance of n happy lives, where $q < p$.

TABLE 2

TWIN EARTH

Probability	p	q	q	q	…	q
Safe	1	1	1	1	…	1
Risky	$n+1$	0	0	0	…	0
Twin Earth	0	1	2	3	…	n

When you take into account the prospect Twin Earth is facing, your options are as follows (see table 3):

Mixed Prospects:

Safe $\oplus$ *Twin Earth* Gives p chance of one happy life, q chance of two happy lives, q chance of three happy lives, q chance of four happy lives, … , q chance of $n+1$ happy lives.

Risky $\oplus$ *Twin Earth* Gives p chance of $n+1$ happy lives, q chance of one happy life, q chance of two happy lives, q chance of three happy lives, … , q chance of n happy lives.

Table 3

Mixed Prospects

Probability	p	q	q	…	q
Safe $\oplus$ *Twin Earth*	1	2	3	…	$n+1$
Risky $\oplus$ *Twin Earth*	$n+1$	1	2	…	n

Next, given that p is greater than q, we may split the first column of table 3 into two columns that give probabilities $p-q$ and q (respectively), as shown in the following table:

Table 4

Mixed Prospects: Split

Probability	$p-q$	q	q	q	…	q
Safe $\oplus$ *Twin Earth*	1	1	2	3	…	$n+1$
Risky $\oplus$ *Twin Earth*	$n+1$	$n+1$	1	2	…	n

Next, we may reorder the outcomes of Risky $\oplus$ Twin Earth that are associated with probability q by moving each of them to the column on their left (see table 5). The leftmost outcome associated with probability q (i.e., $n+1$) is moved to the rightmost column (where n is in table 4).

Table 5

Mixed Prospects: Reorder

Probability	$p-q$	q	q	q	…	q
Safe $\oplus$ *Twin Earth*	1	1	2	3	…	$n+1$
Risky $\oplus$ *Twin Earth*	$n+1$	1	2	3	…	$n+1$

It is now evident from table 5 that the only difference between Safe $\oplus$ Twin Earth and Risky $\oplus$ Twin Earth is that the former gives probability $p - q$ of one happy life, while the latter gives the same probability of $n + 1$ happy lives. As it is better to obtain $n + 1$ happy lives than just one happy life, Risky $\oplus$ Twin Earth stochastically dominates Safe $\oplus$ Twin Earth: For all outcomes, it gives an at least as high probability of an at least as great outcome as Safe $\oplus$ Twin Earth does, and for one outcome, Risky $\oplus$ Twin Earth gives a greater probability of an at least as great outcome as Safe $\oplus$ Twin Earth does. So, by Stochastic Dominance, Risky $\oplus$ Twin Earth is better than Safe $\oplus$ Twin Earth.

Finally, given that Risky $\oplus$ Twin Earth is better than Safe $\oplus$ Twin Earth, it follows by Simple Separability that Risky is better than Safe: Probability Fanaticism is true. So, Probability Fanaticism follows from Simple Separability and Stochastic Dominance.[31] If one wishes to avoid Probability Fanaticism, one must reject Simple Separability or Stochastic Dominance, which are both intuitively compelling.

[31] This argument assumes that the number and not the location of happy lives is all that matters. More generally, Probability Fanaticism follows from Stochastic Dominance, Simple Separability and the following principle:

> **Positive Compensation:** For any near good x and far good y, there is a far good z such that $x \oplus y \sim 0 \oplus z$, and there is a near good w such that $x \oplus y \sim w \oplus 0$.

According to this principle, we can always compensate for making things worse nearby by making things sufficiently better far away (and vice versa). See Russell (2021) for the full argument.

3.4 Stochastic Dominance and Separability are jointly inconsistent

Above we saw how Probability Fanaticism follows from Simple Separability and Stochastic Dominance. However, Stochastic Dominance and a generalization of Simple Separability are jointly inconsistent.[32] This undermines the argument for Probability Fanaticism from Stochastic Dominance and Simple Separability.

Unlike Simple Separability, the generalization of Simple Separability applies to prospects that have an infinite number of possible outcomes. It states the following:[33]

Separability:

i For all near prospects X and Y, and any far prospect Z, $X \succ Y$ if and only if $X \oplus Z \succ Y \oplus Z$.

ii For all far prospects X and Y, and any near prospect Z, $X \succ Y$ if and only if $Z \oplus X \succ Z \oplus Y$.

To see why Stochastic Dominance and Separability are jointly inconsistent, consider the following versions of St. Petersburg games (see table 6):

St. Petersburg Games: A fair coin is flipped until it comes up heads.

St. Petersburg Gives 2^n happy lives, where n is the number of coin flips (and otherwise it gives nothing).

[32] This argument is from Russell (2021).

[33] Russell (2021, p. 15).

St. Petersburg$^-$ Gives $2^n - 1$ happy lives, where n is the number of coin flips (and otherwise it gives nothing).

St. Petersburg$^-$ gives the same probabilities as the St. Petersburg game but slightly worse outcomes. It seems clear that St. Petersburg is better than St. Petersburg$^-$; indeed, this is what Stochastic (and Statewise) Dominance tells us.[34]

Table 6

St. Petersburg Games

No. of flips	1	2	3	...
Probability	$1/2$	$1/4$	$1/8$	...
St. Petersburg	2	4	8	...
St. Petersburg$^-$	1	3	7	...

Separability then tells us that two copies of St. Petersburg, one here and the other in a distant galaxy, are better than two copies of St. Petersburg$^-$, one here and the other in a distant galaxy: As St. Petersburg is better than St. Petersburg$^-$, by Separability, St. Petersburg $\oplus$ St. Petersburg is better than St. Petersburg$^-$ $\oplus$ St. Petersburg. Again, because St. Petersburg is better than St. Petersburg$^-$, St. Petersburg$^-$ $\oplus$ St. Petersburg is better than St. Petersburg$^-$ $\oplus$ St. Petersburg$^-$. Thus, by transitivity, St. Petersburg $\oplus$ St. Petersburg is better than St. Petersburg$^-$ $\oplus$ St. Petersburg$^-$.

However, we can arrange the mechanisms of these games so that St. Petersburg$^-$ $\oplus$ St. Petersburg$^-$ stochastically dominates St. Petersburg $\oplus$ St. Petersburg. In this

[34]According to Statewise Dominance, a prospect is better than another prospect if it gives a better outcome in all states.

case, the results of the two St. Petersburg games depend on the outcome of flipping a dime. And, the results of the two St. Petersburg^{-} games depend on the outcomes of flipping the dime and a penny (see table 7):

Correlated St. Petersburg Games: A dime is flipped until it comes up heads, and a penny is flipped once.

Near and Far St. Petersburg Give 2^n happy lives, where n is the number of coin flips with the dime.

Near St. Petersburg^{-} Gives one happy life if the penny comes up heads. Otherwise, it gives twice as much as St. Petersburg minus one.

Far St. Petersburg^{-} Gives one happy life if the penny comes up tails. Otherwise, it gives twice as much as St. Petersburg minus one.

TABLE 7

CORRELATED ST. PETERSBURG GAMES

Outcome	H, 1	H, 2	H, 3	…	T, 1	T, 2	T, 3	…
Probability	1/4	1/8	1/16	…	1/4	1/8	1/16	…
Near St. Petersburg	2	4	8	…	2	4	8	…
Far St. Petersburg	2	4	8	…	2	4	8	…
Near St. Petersburg^{-}	1	1	1	…	3	7	15	…
Far St. Petersburg^{-}	3	7	15	…	1	1	1	…

'H' and 'T' indicate the outcome of flipping the penny, and '1', '2', … indicate the number of coin flips with the dime.

Note that both the Near and the Far St. Petersburg games give the same probabilities of the same outcomes as the St. Petersburg game in table 6. Similarly, both the Near and the Far St. Petersburg$^-$ games give the same probabilities of the same outcomes as the St. Petersburg$^-$ game in table 6. Thus, Near St. Petersburg $\oplus$ Far St. Petersburg should be better than Near St. Petersburg$^-$ $\oplus$ Far St. Petersburg$^-$.

However, as seen in table 8, Near St. Petersburg $\oplus$ Far St. Petersburg gives the same probabilities of the same outcomes as Near St. Petersburg$^-$ $\oplus$ Far St. Petersburg$^-$. In every state, they result in the same number of happy lives. So, Near St. Petersburg $\oplus$ Far St. Petersburg is stochastically equivalent to Near St. Petersburg$^-$ $\oplus$ Far St. Petersburg$^-$. By Stochastic Dominance, each prospect is at least as good as the other. Therefore, they are equally good.

Table 8

Mixed St. Petersburg Games

Outcome	$H, 1$	$H, 2$	$H, 3$	…	$T, 1$	$T, 2$	$T, 3$	…
Probability	1/4	1/8	1/16	…	1/4	1/8	1/16	…
St. Petersburg $\oplus$ *St. Petersburg*	4	8	16	…	4	8	16	…
St. Petersburg$^-$ $\oplus$ *St. Petersburg*$^-$	4	8	16	…	4	8	16	…

'H' and 'T' indicate the outcome of flipping the penny, and
'1', '2', … indicate the number of coin flips with the dime.
'Near' and 'Far' have been omitted from the prospects' names.

Here is a recap of the argument: Stochastic Dominance tells us that St. Petersburg is better than St. Petersburg$^-$. Separability then tells us that Near St. Petersburg $\oplus$ Far St. Petersburg is better than Near St. Petersburg$^-$ $\oplus$ Far St. Petersburg$^-$. However, Near St. Petersburg $\oplus$ Far St. Petersburg is stochastically equivalent to Near St. Petersburg$^-$ $\oplus$ Far St. Petersburg$^-$. So, they must be equally good. But Near St. Petersburg $\oplus$ Far St. Petersburg cannot both be better than and equally as good as Near St. Petersburg$^-$ $\oplus$ Far St. Petersburg$^-$. Thus, Stochastic Dominance and Separability are jointly inconsistent. Either St. Petersburg is not better than St. Petersburg$^-$, even though it stochastically dominates it. Alternatively, Near St. Petersburg $\oplus$ Far St. Petersburg is not better than Near St. Petersburg$^-$ $\oplus$ Far St. Petersburg$^-$, even though Separability tells us so.

To summarize, Stochastic Dominance and Simple Separability imply Probability Fanaticism. However, Stochastic Dominance and a generalization of Simple Separability are jointly inconsistent.[35] Moreover, whatever the justification of Simple Separability is should also apply to Separability. As Russell (2021, pp. 14–15) writes: "What would the motivation be for it [Simple Separability] that is not also motivation for the unrestricted principle [Separability]? It can't be simply the idea that if what is going on in distant space and time is the same for both of two options, then it is irrelevant to which is better. That idea supports full-fledged Separability. So is there something special about simple prospects that makes their value insensitive to what is going on in distant space and time?" Unless there is some unique justification for Simple Separability that does not also apply to the generalized version, one has no reason to accept Simple Separability if one rejects Separability. And, Stochastic Dominance tells us that Separability is wrong. Thus, for a lack of a unique justification for Simple Separability, the argument for Probability Fanaticism from Stochastic Dominance and Simple Separability does not go through.

[35] As before, this argument assumes that the number, and not the location, of happy lives is all that matters. More generally, Stochastic Dominance, Separability and Positive Compensation are jointly inconsistent. See Russell (2021) for the full argument.

3.5 Stochastic Dominance, Negative Reflection and Background Independence imply Probability Fanaticism

We have seen that Simple Separability and Stochastic Dominance imply Probability Fanaticism, but Separability and Stochastic Dominance are jointly inconsistent. This section shows that Stochastic Dominance, together with two other plausible principles, implies Probability Fanaticism.[36]

Consider the following prospects:

Safe⋆ vs. Risky⋆:

Safe⋆ Certainly gives a good outcome (utility v).

Risky⋆ Gives a tiny probability $p > 0$ of a great outcome (utility V).

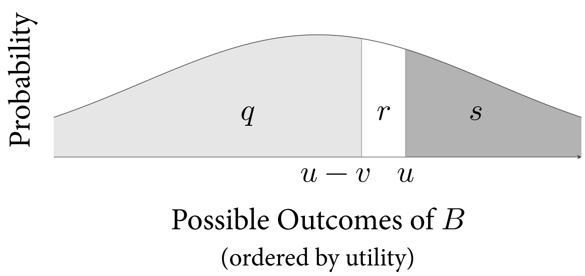
It can be shown that for some background prospect B, which is probabilistically independent of both Safe⋆ and Risky⋆, Risky⋆ $\oplus$ B stochastically dominates Safe⋆ $\oplus$ B.[37] For this to happen, we need Risky⋆ $\oplus$ B to have at least as high a probability as Safe⋆ $\oplus$ B of resulting in at least utility u, for all possible utilities u. Choose any utility $u < V$. Safe⋆ certainly gives utility v, so the probability that Safe⋆ $\oplus$ B gives at least utility u is the probability that B gives at least utility $u - v$ (area $r + s$ in

[36] This argument is from Wilkinson (2022), and the presentation follows closely Russell (2021).

[37] Wilkinson (2022, §VI). See also Tarsney (2020). Tarsney (2020) explores the idea that Stochastic Dominance is a sufficient principle of rationality. Prospects that give a higher expected utility but would not otherwise stochastically dominate their alternatives can become stochastically dominant given sufficient background uncertainty. However, background uncertainty generates stochastic dominance much less readily when the prospect involves tiny probabilities of huge payoffs. So, Tarsney (2020) argues that Stochastic Dominance as a sufficient principle of rationality can vindicate the intuition that we are often permitted to decline gambles like Pascal's Mugging or the St. Petersburg game. However, as the following argument shows, Stochastic Dominance sometimes demands that we make fanatical choices.

the graph below). Risky* $\oplus$ B, in turn, gives at least utility u if either Risky* gives a great outcome (utility $V > u$) or if B gives at least utility u. Denote the probability that B gives at least utility u by s (see area s in the graph below). So, the probability that Risky* $\oplus$ B gives at least u is $p + (1 - p)s$.

PROBABILITY DISTRIBUTION OF B



It can be shown that this (the probability that Risky* $\oplus$ B gives at least u) will be greater than the probability that Safe* $\oplus$ B gives at least u if the area r is less than or equal to the area q multiplied by p.[38] So, if the area r is small enough compared to area q, then Risky* $\oplus$ B gives an at least as high probability of u as Safe* $\oplus$ B does. For this to happen with all u, the interval between u and $u - v$ needs to be tiny enough. And for that to happen, the probabilities in B must go down slowly

[38]

$$
\begin{aligned}
&p + s(1-p) \geq r + s \\
\iff\ & s + p(1-s) \geq r + s \\
\iff\ & p(q+r) \geq r \\
\iff\ & r \leq pq + pr \\
\iff\ & r(1-p) \leq pq \\
\iff\ & r \leq \frac{pq}{1-p}, \text{ for which } r \leq pq \text{ is a sufficient condition.}
\end{aligned}
$$

enough as we approach $-\infty$ and rise and fall quickly enough as we pass the peak of the curve. There are some probability distributions with this property.[39] So, with some background prospect B, Risky* $\oplus$ B is better than Safe* $\oplus$ B by Stochastic Dominance.

Next, consider the following principle:[40]

> **Negative Reflection:** For prospects X and Y and a question Q, if X is not better than Y conditional on any possible answer to Q, then X is not better than Y unconditionally.

Given that Risky* $\oplus$ B is better than Safe* $\oplus$ B (by Stochastic Dominance), Negative Reflection tells that B must have some possible outcome b such that Risky* $\oplus$ b is better than Safe* $\oplus$ b.

Finally, consider the following principle:[41]

> **Background Independence:** For any near prospects X and Y and any far outcome a, $X \succ Y$ if and only if $X \oplus a \succ Y \oplus a$.

---

[39] Wilkinson (2022, §VI) and Tarsney (2020).

[40] Russell (2021, p. 19). Compare *Negative Reflection* to the following related principle (Russell, 2021, p. 23):

> **Positive Reflection:** For prospects X and Y and a question Q, if X is at least as good as Y conditional on any possible answer to Q, then X is at least as good as Y unconditionally.

Both reflection principles are related to the Sure Thing Principle due to Savage (1972, pp. 21–22):

> **The Sure Thing Principle:** For all prospects X and Y, if an agent would not prefer X over Y if they learnt that some event E has happened, or if they learnt that E has not happened, then the agent does not prefer X over Y. Moreover, if the agent would prefer Y to X if they learnt that E has happened, and they would not prefer X to Y if they learnt that E has not happened, then the agent prefers Y to X.

[41] Wilkinson (2022, p. 467) and Russell (2021, p. 28).

Background Independence is similar to Separability. But unlike Separability, it requires that the 'background prospect' involves no uncertainty.[42] Given that Risky* $\oplus$ b is better than Safe* $\oplus$ b (by Stochastic Dominance and Negative Reflection), Background Independence implies that Risky* is better than Safe*. This means that Probability Fanaticism is true.

To conclude, three plausible principles, Stochastic Dominance, Negative Reflection and Background Independence, imply Probability Fanaticism. Thus, to deny Probability Fanaticism, one must reject Stochastic Dominance, Negative Reflection or Background Independence.

3.6 Stochastic Dominance and Negative Reflection imply Probability Fanaticism is false

We just saw how Stochastic Dominance, Negative Reflection and Background Independence imply Probability Fanaticism. However, if Probability Fanaticism is true, then two of the premises of the previous argument—Stochastic Dominance and Negative Reflection—are jointly inconsistent.[43] Thus, the argument cannot be sound.

To see why Stochastic Dominance and Negative Reflection are jointly inconsistent if Probability Fanaticism is true, consider the following versions of the St. Petersburg game:

[42] Background Independence is related to the Egyptology objection to the Average View in population ethics. See McMahan (1981, p. 115) and Parfit (1984, p. 420).

[43] This argument is from Russell (2021, §3) and Russell and Isaacs (2021). Also see Chalmers (2002) and Beckstead and Thomas (2020, §4).

St. Petersburg Games: A fair coin is flipped until it comes up heads.

St. Petersburg Gives 2^n happy lives, where n is the number of coin flips.

St. Petersburg$^+$ Gives $2^n + 1$ happy lives, where n is the number of coin flips.

The outcomes of St. Petersburg$^+$ are better than the outcomes of St. Petersburg. So, by Stochastic Dominance, St. Petersburg$^+$ is better than St. Petersburg.

However, if Probability Fanaticism is true, then none of the *outcomes* of St. Petersburg$^+$ are as good as the *prospect* St. Petersburg. This is because St. Petersburg and St. Petersburg$^+$ are better than any possible finite payoffs. So, any possible payoff of St. Petersburg$^+$ is worse than the *prospect* St. Petersburg. Negative Reflection, therefore, implies that St. Petersburg$^+$ is not better than St. Petersburg. Conditional on any way St. Petersburg$^+$ could turn out, St. Petersburg$^+$ is not better than St. Petersburg, so St. Petersburg$^+$ cannot be better than St. Petersburg. However, St. Petersburg$^+$ is better than St. Petersburg by Stochastic Dominance. So, if Probability Fanaticism is true, either Stochastic Dominance or Negative Reflection needs to go. They are jointly inconsistent. Thus, Stochastic Dominance and Negative Reflection imply that Probability Fanaticism is false.[44]

Suppose that Probability Fanaticism keeps Stochastic Dominance (and gives up Negative Reflection). In that case, it is dynamically inconsistent and vulnera-

[44] Stochastic Dominance and Negative Reflection imply that Probability Fanaticism is false. They cannot, therefore, be used in an argument for Probability Fanaticism. However, a principle that is related to Negative Reflection, together with Stochastic Dominance and Background Independence, implies that Positive Fanaticism or Negative Fanaticism is true. See Russell (2021, pp. 37–38).

ble to money pumps.[45] Consider for example the following case: You start with St. Petersburg$^+$. Once the result of St. Petersburg$^+$ is known, you can pay \$100 to exchange the outcome St. Petersburg$^+$ for the prospect St. Petersburg. Because any possible finite payoff of St. Petersburg$^+$ is worse than the prospect St. Petersburg, you would accept this trade. But before finding out the result of St. Petersburg$^+$, you can pay \$50 to simply keep the prospect St. Petersburg$^+$ and receive no further offers. You know that if you do not pay this \$50, you will end up with prospect St. Petersburg and with \$100 less in your wallet. But if you do pay this \$50, then you will end up with prospect St. Petersburg$^+$, and you will only have paid \$50. Therefore, by Stochastic Dominance, you should pay \$50 to avoid any further offers. But you have then been money pumped, as you have paid for something that you could have kept for free had you refused all the offers.[46] So, Probability Fanaticism combined with Stochastic Dominance is vulnerable to money pumps.

To conclude, this section has discussed arguments for and against Probability Fanaticism. §3.1 showed that Anti-Timidity and transitivity imply Probability Fanaticism. Then, §3.2 showed that More is Better and Simple Separability imply Probability Fanaticism. §3.3 showed that Stochastic Dominance and Simple Separability imply Probability Fanaticism. However, §3.4 showed that Stochastic

[45] This argument is from Russell and Isaacs (2021, p. 4 n. 5). Russell and Isaacs (2021) show that Probability Fanaticism violates Countable Independence, which is similar to Negative Reflection.

[46] If Probability Fanaticism rejects both Stochastic Dominance and Negative Reflection, then St. Petersburg$^+$ is not better than St. Petersburg. However, you are still permitted to pay to keep St. Petersburg$^+$ and receive no further offers. In fact, you are permitted to pay any finite amount to receive no further offers; whichever finite sum you pay, you will face a prospect with infinite expected utility. Thus, Probability Fanaticism still permits you to make a series of trades that results in a sure loss.

Dominance and (generalized) Separability are jointly inconsistent. §3.5 showed that Stochastic Dominance, Negative Reflection and Background Independence imply Probability Fanaticism. But §3.6 showed that Stochastic Dominance and Negative Reflection imply that Probability Fanaticism is false. §3.6 also showed that Probability Fanaticism is vulnerable to exploitation by money pumps. The debate between proponents and opponents of Probability Fanaticism is inconclusive, as there are strong arguments for and against it. However, as Russell (2021, p. 5) writes, "Whatever the truth of the matter, the ethics of huge numbers is deeply weird and full of surprises."

4 Bounded utilities

The rest of the sections discuss alternatives to Probability Fanaticism. This section explores the idea that utilities are bounded above and below.

Boundedness of utilities has been discussed as a possible alternative to Probability Fanaticism.[47] If utilities are real valued, then Boundedness means the following:

> **Boundedness:** There is some $M \in \mathbb{R}$ such that for all outcomes x, $|u(x)| < M$.

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

The following discussion focuses on Boundedness in the context of Expected Utility Theory. As discussed in Chapter 1 of this thesis, standard axiomatizations

[47] See Beckstead and Thomas (2020, §2.1) and Chapter 1 of this thesis.

of expected-utility maximization require utilities to be bounded.[48] Bounded utilities are, therefore, the standard in decision theory. However, bounded utilities seem troubling from the point of view of ethics. It seems odd that, for example, additional happy lives matter less the more happy lives there already are or that additional headaches matter less the more headaches (or other negative experiences) there already are. Also, bounded utilities imply that it is better to save some (very large) number n of lives for sure than to save *any number* of lives with a probability of almost one.[49] This happens when the value of n happy lives is close to the upper bound of utilities as then additional happy lives do not contribute much to expected utility.

Boundedness gives ethically even more untenable prescriptions. Consider for example the following prospects (see table 9):

> **Happy Lives vs. Headaches:** A fair coin is flipped.
>
> *Prospect A* Gives some large number m of happy lives with heads, and one person gets a headache with tails.
>
> *Prospect B* Gives some much larger number M of happy lives with heads, and two people get headaches with tails.

Suppose that the values of m and M happy lives are close to the upper bound of utilities. In that case, the additional happy lives in B may not contribute enough to B's expected utility to outweigh the disvalue of the possible additional headache.

[48] See Kreps (1988, pp. 63–64), Fishburn (1970, pp. 194, 206–207), Hammond (1998, pp. 186–191) and Russell and Isaacs (2021).

[49] More generally, Boundedness violates Anti-Timidity. See Beckstead and Thomas (2020, §2.1).

Then, Boundedness (from above) implies that A is better than B—which seems wrong from an ethical point of view.[50]

TABLE 9

HAPPY LIVES VS. HEADACHES

	Heads	**Tails**
A	Many happy lives	One headache
B	Very many happy lives	Two headaches

Next, consider the following prospects (see table 10):

Unhappy Lives vs. Lollipops: A fair coin is flipped.

Prospect C Gives some large number m of unhappy lives with heads, and one person gets a lollipop with tails.

Prospect D Gives a much larger number M of unhappy lives with heads, and two people get lollipops with tails.

For similar reasons as explained above, D may be better than C if utilities are bounded below.[51] The implications of Boundedness is ethically untenable; the possibility of one additional lollipop should not compensate for an equally likely chance of many additional unhappy lives.

[50] This argument is from Beckstead and Thomas (2020, §3.3).

[51] See Beckstead and Thomas (2020, §3.4).

Table 10
Unhappy Lives vs. Lollipops

	Heads	Tails
C	Many unhappy lives	One lollipop
D	Very many unhappy lives	Two lollipops

Boundedness also implies that sometimes one should choose a small probability of a mediocre payoff instead of a high probability of a great payoff—which violates More is Better. To see how this violation happens, consider the following prospects (see table 11):[52]

Great vs. Mediocre Past:

Great Gives some great payoff (such as very many happy lives) if humanity's past was great (high probability p); otherwise, nothing happens.

Mediocre Gives some mediocre payoff (such as a few happy lives) if humanity's past was mediocre (small probability $1 - p$); otherwise, nothing happens.

In this case, Boundedness implies that Mediocre might be better than Great. If humanity's past was great (in which case the value of the world is near the upper bound of utilities), then the great payoff does not contribute much to utility. However, if humanity's past was mediocre, then the mediocre payoff makes a large contribution to utility. Thus, Boundedness implies that one should choose a small

[52]This argument is from Beckstead and Thomas (2020, §3.5).

probability of a mediocre payoff (and otherwise nothing) instead of a high probability of a great payoff (and otherwise nothing). This is a violation of More is Better.

Table 11

Great vs. Mediocre

	Great past	**Mediocre past**
Probability	p	$1 - p < p$
Great	Many happy lives	Nothing
Mediocre	Nothing	A few happy lives

We have seen that Boundedness has ethically worrying implications. Chapter 1 of this thesis shows another troubling feature of Boundedness. It shows that decision theories on which utilities are bounded, such as Expected Utility Theory, violate Ex Ante Pareto if combined with an additive axiology, such as Total Utilitarianism. According to Total Utilitarianism, a population is better than another just in case the total quantity of well-being it contains is greater. Ex Ante Pareto, in turn, states the following:

> **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.

The combination of Expected Utility Theory and Total Utilitarianism violates Ex Ante Pareto 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. This non-linear transformation is required if one has a non-zero credence in the possibility that an infinite or arbitrarily large number of individuals exist. But it is also required if one wishes to avoid Probability Fanaticism. 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. Chapter 1 also discusses how this relates to a well-known result in this area, namely, Harsanyi's social aggregation theorem.

Chapter 2 of this thesis is somewhat related to the discussion of Boundedness. It points out that standard axiomatizations of Expected Utility Theory violate Statewise Dominance with prospects that involve possible states of zero probability. Statewise Dominance says the following:

> **Statewise Dominance:** If the outcome of prospect X is at least as good as the outcome of prospect Y in all states, and the outcome of X is better than the outcome of Y in some possible state, then X is better than Y.

At least at first glance, Expected Utility Theory tells us to be indifferent between two prospects when they are otherwise the same, except that one gives a better outcome than the other in a possible state of zero probability. But as some have suggested, Expected Utility Theory might be supplemented with dominance reasoning to get the verdict that the dominating prospect is better than the dominated one. However, Chapter 2 shows that if Expected Utility Theory is supplemented with dominance reasoning in this way, it will violate the Continuity axiom of Expected Utility

Theory. So, if an expected-utility maximizer wishes to retain Statewise Dominance even in cases that involve possible states of zero probability, they must adopt some axiomatization of Expected Utility Theory that does not have Continuity as one of the axioms.

To conclude, bounded utilities have been proposed as an alternative to Probability Fanaticism. Boundedness follows from standard axiomatizations of Expected Utility Theory, so it is the orthodox view in decision theory. However, Boundedness is troubling from an ethical point of view. For example, if utilities are bounded, it is better to save some (very large) number n of lives for sure than to save *any number* of lives with a probability of almost one. Also, it sometimes implies that the possibility of a very large number of additional happy lives cannot compensate for the disvalue of an equally likely additional headache. Similarly, it sometimes implies that the possibility of a single additional lollipop can compensate for the disvalue of an equally likely possibility of a very large number of unhappy lives. Also, Boundedness sometimes implies that one should choose a small probability of a mediocre outcome over a high probability of a great outcome. Furthermore, Chapter 1 of this thesis shows that decision theories on which utilities are bounded violate Ex Ante Pareto if combined with an additive axiology. Also, as shown in Chapter 2, standard axiomatizations of Expected Utility Theory violate Statewise Dominance in cases that involve possible states of zero probability.

5 Probability Discounting

This section discusses another alternative to Probability Fanaticism: discounting small probabilities.[53]

5.1 Discounting small probabilities

In response to cases that involve very small probabilities of huge payoffs, some have argued that we should discount very small probabilities down to zero—let's call this *Probability Discounting*. For example, Monton (2019) argues that small probabilities should be discounted down to zero, while Smith (2014*b*) argues that one is rationally permitted, but not required, to do so.[54] Probability Discounting avoids the counterintuitive implication that you should pay the stranger in Pascal's Mugging because it tells you to discount the tiny probability of the mugger telling the truth. Similarly, Probability Discounting allows one to value the St. Petersburg game at a reasonable price. In fact, Probability Discounting was originally proposed by Nicolaus Bernoulli as a solution to the St. Petersburg paradox.[55] He

[53]Note that, unlike here, 'discounting' typically does not mean ignoring altogether or bringing all the way down to zero. For example, 'temporal discounting' does not typically mean disvaluing positive outcomes in the future altogether, but instead, holding them less valuable than similar outcomes in the present.

[54]Smith argues that discounting small probabilities allows one to get a reasonable expected utility for the Pasadena game (see [Nover and Hájek 2004]). See Hájek (2014), Isaacs (2016) and Lundgren and Stefánsson (2020) for criticisms of discounting small probabilities. There is a related discussion on *de minimis* principles, on which a risk can be ignored or treated very differently from other risks if the risk is sufficiently small. See for example Peterson (2002) and Lundgren and Stefánsson (2020).

[55]Monton (2019) calls discounting small probabilities 'Nicolausian discounting' after Nicolaus Bernoulli. Other proponents of Probability Discounting include, for example, Buffon and Condorcet. See Hey et al. (2010) and Monton (2019, pp. 16–17).

writes: "[T]he cases which have a very small probability must be neglected and counted for nulls, although they can give a very great expectation."[56]

There are many ways of cashing out Probability Discounting. On one of the simplest versions of this view (i.e., *Naive Discounting*), one should conditionalize on very-small-probability outcomes not occurring and then maximize expected utility. On this view, there is some threshold probability such that outcomes whose probabilities are below this threshold are ignored. A slightly more complicated version (i.e., *Lexical Discounting*) uses very-small-probability outcomes as tiebreakers in cases where the prospects would otherwise be equally good. Both of these versions ignore *outcomes* associated with tiny probabilities. Instead, one could ignore *states* of the world that have tiny probabilities of occurring (as *State Discounting* does). Chapter 4 of this thesis discusses these and other versions of Probability Discounting in more detail. It explores what the most plausible version of Probability Discounting might look like and what are some problems such theories face.

5.2 Implications of Probability Discounting

Two chapters of this thesis examine the implications of Probability Discounting for population ethics and the value of the far future. These implications are briefly outlined below.

POPULATION ETHICS. The Repugnant Conclusion, introduced by Parfit, states:[57]

> "For any possible population of at least ten billion people, all with a

[56] Pulskamp (2013, p. 2).
[57] Parfit (1984, p. 388).

> 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."

The Repugnant Conclusion is a consequence of standard Total Utilitarianism. The Repugnant Conclusion strikes many as an unacceptable consequence, and various attempts at constructing an alternative population axiology to Total Utilitarianism have been made.[58] Nebel (2019) argues for the Repugnant Conclusion via the "Intrapersonal Repugnant Conclusion", on which certainty of a mediocre life is better for individuals than a sufficiently small chance of an excellent life. In Chapter 3 of this thesis, I deny that acceptance of the Intrapersonal Repugnant Conclusion leads us to the Repugnant Conclusion. I point out that on many views which avoid the Repugnant Conclusion, we should discount small probabilities down to zero to avoid an implausibly reckless decision theory. If we do, then Nebel's crucial premise of Ex Ante Pareto fails because discounting at the individual level can fail to match up with discounting at the population level. Thus, Probability Discounting helps us avoid the Repugnant Conclusion.

Value of the Far Future. Chapter 6 of this thesis discusses the implication of Probability Discounting for

> **Longtermism:** In the most important decision situations, our acts' expected influence on the value of the world is mainly determined by their possible consequences in the far future.[59]

[58] For an overview, see Greaves (2017).
[59] MacAskill (2019) and Greaves and MacAskill (2021).

According to Longtermism, morally speaking what matters the most is the far future. The case for Longtermism is straightforward: Given the enormous number of people who might exist in the far future, even a tiny probability of affecting how the far future goes outweighs the importance of our acts' consequences in the near term. But if we discount very small probabilities down to zero, we may have an objection to Longtermism provided that its truth depends on tiny probabilities of vast value. Contrary to this, Chapter 6 argues that discounting small probabilities does not undermine Longtermism. However, Probability Discounting might have implications for what longtermists should focus on.[60]

5.3 Problems with Probability Discounting

Probability Discounting might allow us to reject Probability Fanaticism and escape the Repugnant Conclusion. But it also faces some serious problems, as outlined below.

THRESHOLD. One obvious problem with Probability Discounting is where the 'discounting threshold' is located. When are probabilities small enough to be discounted? Some have proposed possible thresholds. For example, Buffon suggested

[60] For example, Probability Fanaticism might imply that 'effective altruists' should accept Pascal's Wager. See footnote 4. They would have then made a full circle: Donate 10% of your income to your local church, mosque or synagogue. Or Probability Fanaticism might imply something weirder (see for example Wilkinson [2022, pp. 445–446]). In contrast, Probability Discounting allows one to escape this implication, provided that one's credence in heaven is low enough. Beckstead and Thomas (2020, §5) show that Probability Fanaticism leads to *Infinity Obsession*:

> **Infinity Obsession:** Any non-zero probability, no matter how small, of an infinite payoff is better than any finite payoff for sure.

that the threshold should be one-in-ten-thousand. And Condorcet gave an amusingly specific threshold: 1 in 144,768. Buffon chose his threshold because it was the probability of a 56-year-old man dying in one day—an outcome reasonable people usually ignore.[61] Condorcet had a similar justification.[62] More recently, Monton (2019, p. 17) has suggested a threshold of 1 in 2 quadrillion—significantly lower than the thresholds given by the historical thinkers. Monton (2019, §6.1) thinks that the threshold is subjective within reason: There is no single objective threshold for everybody.

However, there seems to be no way of choosing the discounting threshold such that Probability Discounting rules out all and only the objectionable choices.[63] For example, suppose the discounting threshold is just below 1 in 2 quadrillion. In that case, a prospect that gives any finite payoff for sure, no matter how good, is worse than a 1 in 2 quadrillion probability of some other finite payoff (assuming unbounded utilities). But a prospect with a 1 in 2 quadrillion probability does not seem less objectionable than a prospect with a slightly lower probability. So, Probability Discounting does not solve the problem it was meant to solve, as it still implies objectionably fanatical choices. However, this problem might be somewhat mitigated by letting the discounting threshold be vague.[64]

Individuation Problem. Another problem with Probability Discounting comes

[61] Hey et al. (2010, p. 257). See Monton (2019, pp. 8–9) for a discussion of Buffon's view.

[62] Condorcet's justification for his threshold is that 1 in 144,768 was the difference between the probability that a 47-year-old man would die within 24 hours and the probability that a 37-year-old man would, and that difference would not keep anyone awake at night. See Monton (2019, pp. 16–17).

[63] This point is raised by Beckstead and Thomas (2020, §3.5).

[64] Beckstead and Thomas (2020, p. 20). See also Lundgren and Stefánsson (2020, p. 911).

from individuating outcomes and states. The problem is that if we individuate outcomes/states very finely by giving a great deal of information about them, then all outcomes/states will have probabilities below the threshold. As discussed later in this thesis, one possible solution is to individuate outcomes by utilities. The idea is that outcomes/states are considered the "same" outcome/state if their associated utilities are the same.

Dominance Violations. One problem for some versions of Probability Discounting is that they violate dominance.[65] Imagine a lottery that gives you a tiny probability of some prize (and otherwise nothing), and compare this to a lottery that surely gives you nothing. The former lottery dominates the latter, but some versions of Probability Discounting say they are equally good. One can solve this dominance violation by considering very-small-probability outcomes/states as tiebreakers in cases where the prospects are otherwise equally good. However, this is not enough to avoid violating dominance because the resulting views still violate dominance in more complicated cases (as discussed in Chapter 4).

Money Pumps. Some versions of Probability Discounting, such as *Tail Discounting*, avoid the above mentioned dominance violations. According to Tail Discounting, one should first order all the possible outcomes of a prospect in terms of betterness. Then one should ignore the 'tails', that is, the very best and the very worst outcomes. Tail Discounting solves the problems with individuating outcomes and dominance violations. But it also has one big problem: It can be money pumped

[65] Isaacs (2016), Smith (2016), Monton (2019, pp. 20–21), Lundgren and Stefánsson (2020, pp. 912–914) and Beckstead and Thomas (2020, §2.3) also discuss Probability Discounting and dominance violations.

(as discussed in Chapter 4). So, someone with this view would end up paying for something they could have kept for free, which makes Tail Discounting less plausible as a theory of instrumental rationality.

In fact, vulnerability to exploitation by money pumps may be one of the most challenging problem for all versions of Probability Discounting. One money pump, in particular, presents a difficult challenge. Probability discounters are vulnerable to this money pump as a result of violating the Independence axiom of Expected Utility Theory. The basic problem for Probability Discounting is that by mixing gambles, one can arbitrarily reduce the probabilities of different states or outcomes within the compound lottery until these probabilities end up below the discounting threshold. Therefore, mixtures of gambles can end up being valued differently than the gambles that are mixed together. How probability discounters can avoid exploitation by money pumps is discussed in Chapter 4 and, in more detail, in Chapter 5.

Ex Ante Pareto. As discussed in Chapter 3, accepting Ex Ante Pareto and engaging in Probability Discounting gets one in trouble. Consider, for example, the following case:

> **Celebratory Gunfire:** Someone shoots into the air in an area full of people during a celebration, which causes people to feel excitement for a few seconds. The probability of any particular individual being hit by the bullet when it falls is negligibly small, but there is a high probability that someone is hit by it.

We may suppose that the value of everyone feeling excitement is not enough to outweigh the badness of the likely injury. However, the prospect of shooting into the air is *ex ante* better than not shooting for everyone; each individual feels excitement, and the probability of being hit by the bullet is rationally negligible. Thus, Ex Ante Pareto tells us that shooting into the air is right, even though the bullet will almost certainly hit someone. So, if one accepts Probability Discounting, one should reject Ex Ante Pareto, or one would permit the infliction of arbitrarily severe harms for little or no benefits.

EACH-WE DILEMMAS. Another problem Probability Discounting faces is Each-We Dilemmas, which will be discussed in Chapter 6. According to Parfit (1984, p. 91), a theory faces Each-We Dilemmas if "there might be cases where, if each does better in this theory's terms, we do worse, and vice versa." Each-We Dilemmas arise for Probability Discounting for the same reason as violations of Ex Ante Pareto arise: Probabilities can accumulate. If many individuals discount a tiny probability of some event happening, and the probabilities are sufficiently independent for the different agents, then the total discounted probability can be high. This can result in catastrophic outcomes. Consider, for example, the following case:

> **Asteroid:** An asteroid is heading toward the Earth and will almost certainly hit unless stopped. There are multiple asteroid defense systems, and (unrealistically) each has a tiny probability of hitting the asteroid and preventing a catastrophe. However, the probability that one of them succeeds is high if enough of them try. Attempting to stop

the asteroid involves some small cost ϵ.

If agents discount the probability of them successfully stopping the asteroid and consequently do nothing, then the asteroid will almost certainly hit the Earth. But this outcome could be prevented almost certainly if enough agents attempt to do so. To solve these kind of cases, probability-discounting agents would need to somehow take into account the choices other people face and consider whether the collective has a non-negligible chance of making a difference. However, this solution leads us to another problem: violations of Separability.

Separability. Probability Discounting violates Separability if the choices other people face can affect what you ought to do, even when the other agents are far away and you cannot influence what goes on near them. The solution to Each-We Dilemmas asks us to change our actions depending on what choices other agents face. For example, if there was only a single asteroid defense system, then Probability Discounting would recommend that the agent operating it not attempt to stop the asteroid. However, if there are multiple asteroid defense systems, then this approach would recommend attempting to stop the asteroid because the probability that someone successfully stops it is non-negligible.

Earlier it was shown that Stochastic Dominance and Separability are jointly inconsistent. In Russell's (2021, pp. 13–14) words: "This looks like very bad news for Separability." Since violating Separability is a problem for all theories (on pain of violating Stochastic Dominance), violating Separability may not seem especially worrying for Probability Discounting. However, it was only shown that Stochastic Dominance and Separability are inconsistent in a case where the outcomes (of the near and far prospects) are correlated. In contrast, Probability Discounting violates Separability even when the outcomes are probabilistically independent for the different agents. We might think that probabilistic independence makes violating Separability even worse.

To summarize, I have discussed some problems Probability Discounting faces. These include choosing the discounting threshold, individuating outcomes/states, violating dominance, vulnerability to money pumps, violating Ex Ante Pareto, facing Each-We Dilemmas and violating Separability. These problems will be discussed in more detail in the following chapters of this thesis. The next section discusses some alternative approaches to tiny probabilities of vast value.

6 Alternatives

This section discusses other approaches suggested in response to cases that involve tiny probabilities of huge payoffs.

6.1 Conditionalizing on knowledge

One possibility is to conditionalize on one's knowledge before maximizing expected utility—let's call this *Knowledge-Based Discounting*.[66] It might be argued that, in Pascal's Mugging, you *know* that the mugger will not deliver a thousand quadrillion happy days in the Seventh Dimension. And, possibly, you also know that you will

[66] See Hong (n.d.) and Francis and Kosonen (n.d.) on Knowledge-Based Discounting.

not gain a great payoff with the St. Petersburg game.[67] Thus, conditionalizing on knowledge before maximizing expected utility could solve at least some cases with tiny probabilities of huge payoffs.

But Knowledge-Based Discounting is vulnerable to some of the same problems Probability Discounting faces, such as money pumps.[68] Consider, for example, the following lotteries:

> *Ticket A* Gives a great payoff if you guess all seven lottery numbers correctly (and otherwise it gives nothing).
>
> *Ticket B* Gives a modest positive payoff if you guess at least five lottery numbers correctly (and otherwise it gives nothing).

Suppose you know that ticket A wins nothing, but you do not know that ticket B wins nothing. If it is possible to have knowledge in lottery cases, then there must be some (possibly vague and context-dependent) threshold probability for when a probability is high enough to count as knowledge. We may suppose that the probability of not winning with A is above this threshold, but the probability of not winning with B is below this threshold. Consequently, B is worth some positive amount, while A is worthless (or at most better than nothing). The setup is as follows: You currently have B. If you guess at least five lottery numbers right,

[67] This claim is more contested. It is often argued that one cannot have knowledge in lottery cases, as it seems that one does not know that one's lottery ticket will not win, even though it is very unlikely to win. For a discussion of lottery cases, see for example Smith (2014*a*). See Hong (n.d.) for a defense of Knowledge-Based Discounting in the context of the St. Petersburg paradox. If Knowledge-Based Discounting is to avoid Probability Fanaticism in all cases, then it must be possible to have knowledge in lottery cases, such as the St. Petersburg paradox.

[68] See Francis and Kosonen (n.d.).

then you will be offered A in exchange for B. But if you learn that you guessed five lottery numbers right, you no longer know that you did not guess all seven numbers right. In that case, you would only need to have guessed two more numbers right, and for all you know, you might have. So, you then would prefer A to B and happily accept the trade.

This is unfortunate. Right now, you know that you will not win anything with A. So it would be better to keep B. However, you also know that if you win anything with B, you will accept the trade and end up with A. Luckily, you are offered a chance to avoid this situation: If you pay some amount, you will not be offered A in exchange for B in case you guess at least five numbers right. And, given that B is worth some positive amount while A is worth nothing, you accept this offer. But you have then paid for something you could have kept for free.

More generally, Knowledge-Based Discounting gets you in trouble if there can be cases where you know that P, but some evidence would make you lose the knowledge that P and you do not know that such evidence will not arise.[69] In this case, although you know that A wins nothing, this belief loses the status of knowledge if you guess at least five lottery numbers right. And you do not know that you will not guess at least five lottery numbers right.

To summarize, Knowledge-Based Discounting advises one to conditionalize

[69] Knowledge-Based Discounting might escape this problem if one accepts the *KK Principle*: If one knows that P, then one also knows that one knows it. If the KK principle is true, then either you do not know that you will not guess seven numbers correctly (so A is worth some positive amount), or you know that you will not guess at least five numbers correctly (so neither A nor B is worth any positive amount). But you cannot know that you will not guess seven numbers correctly and be uncertain about whether you might lose this knowledge.

on one's knowledge before maximizing expected utility. Similarly to Probability Fanaticism and Probability Discounting, Knowledge-Based Discounting is diachronically inconsistent and thus vulnerable to money pumps.

6.2 Assigning zero probability

It seems that every approach to tiny probabilities of huge payoffs has serious shortcomings. In order to escape the paradoxes with non-simple lotteries, one might be tempted to assign a zero probability to the possibility of the St. Petersburg game (and its variants).[70] The idea is that one can accept Expected Utility Theory and escape the paradoxical results discussed earlier in this chapter, such as money pumps.

However, this solution seems *ad hoc*. Assignments of probability should only respond to *epistemic reasons*. They should not respond to instrumental reasons, such as getting money pumped.[71] Yet, arguments for Bayesianism often rely on such instrumental reasons: Unless one uses conditionalization to update credences, one will get Dutch Booked. However, something else might be going on in these arguments. Succumbing to a Dutch Book is an indication that one's beliefs about the world are inconsistent. So, the argument for conditionalization is not that failing to conditionalize gets one Dutch Booked, and that is instrumentally bad. Instead,

[70] Various people have suggested this (personal correspondence). Note that this proposal does not avoid Probability Fanaticism—its only purpose is to make Expected Utility Theory behave well with non-simple lotteries. But it says nothing about cases such as Pascal's Mugging.

[71] This is controversial. For example, in epistemology, there is a view that rejects the claim that only epistemic reasons should influence beliefs:

> **Pragmatic Encroachment:** A difference in pragmatic circumstances can constitute a difference in knowledge.

See Ichikawa and Steup (2018, §12).

the argument is that getting Dutch Booked is a symptom of having inconsistent beliefs.[72] So, Dutch Book arguments need not rely on the idea that beliefs (or credence assignments) should respond to instrumental reasons. However, similarly, one might insist that getting money pumped because one accepts Probability Fanaticism is a symptom of having inconsistent beliefs. The money pump shows that there is something wrong with St. Petersburg-style probability and utility assignments. But it is hard to see why that would be the case.

7 Conclusion

Cases that involve tiny probabilities of vast value present a puzzle as it seems that all approaches have implausible implications. The main approaches discussed were Probability Fanaticism, Boundedness and Probability Discounting. First, the chapter discussed two arguments for maximizing expected utility: the long-run argument and representation theorems. Next, it explored Probability Fanaticism, on which tiny probabilities of large positive or negative payoffs can have enormous positive or negative expected utility (respectively). We saw that there are strong arguments for and against Probability Fanaticism. Then, the chapter discussed the possibility that utilities are bounded. Boundedness will be discussed in more de-

[72] This is, in fact, what Lewis (1999, pp. 404–405) argues: "Note also that the point of any Dutch book argument is not that it would be imprudent to run the risk that some sneaky Dutchman will come and drain your pockets. After all, there aren't so many sneaky Dutchmen around; and anyway, if ever you see one coming, you can refuse to do business with him. Rather, the point is that if you are vulnerable to a Dutch book, whether synchronic or diachronic, that means that you have two contradictory opinions about the expected value of the very same transaction. To hold contradictory opinions may or may not be risky, but it is in any case irrational."

tail in Chapters 1 and 2 of this thesis. Finally, the chapter investigated Probability Discounting, on which tiny probabilities should be ignored in practical decision-making. Probability Discounting will be the focus of Chapters 3–6. Some other approaches were also discussed briefly. To conclude, paradoxes concerning tiny probabilities of vast value show that some intuitively compelling principles of rationality must be given up.

References

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. and Thomas, T. (2020), 'A paradox for tiny probabilities and enormous values'. Global Priorities Institute Working Paper No. 10–2020. **URL:** *https://globalprioritiesinstitute.org/nick-beckstead-and-teruji-thomas-a-paradox-for-tiny-probabilities-and-enormous-values/*

Bernoulli, D. (1954), 'Exposition of a new theory on the measurement of risk', *Econometrica* **22**(1), 23–36.

Bostrom, N. (2009), 'Pascal's Mugging', *Analysis* **69**(3), 443–445.

Briggs, R. A. (2019), Normative Theories of Rational Choice: Expected Utility, *in*

E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', Fall 2019 edn, Metaphysics Research Lab, Stanford University.

Buchak, L. (2013), *Risk and Rationality*, Oxford University Press, Oxford.

Chalmers, D. J. (2002), 'The St. Petersburg two-envelope paradox', *Analysis* **62**(2), 155–157.

Fishburn, P. C. (1970), *Utility Theory for Decision Making*, Wiley, New York.

Francis, T. and Kosonen, P. (n.d.), 'Ignore outlandish possibilities'. Unpublished manuscript.

Goodsell, Z. (2021), 'A St Petersburg paradox for risky welfare aggregation', *Analysis* **81**(3), 420–426.

Greaves, H. (2017), 'Population axiology', *Philosophy Compass* **12**(11), e12442.

Greaves, H. and MacAskill, W. (2021), 'The case for strong longtermism'. Global Priorities Institute Working Paper No. 5–2021.
URL: *https://globalprioritiesinstitute.org/hilary-greaves-william-macaskill-the-case-for-strong-longtermism-2/*

Gustafsson, J. E. (forthcoming), *Money-Pump Arguments*, Cambridge University Press, Cambridge.

Hadar, J. and Russell, W. R. (1969), 'Rules for ordering uncertain prospects', *The American Economic Review* **59**(1), 25–34.

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 and 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*

Hey, J. D., Neugebauer, T. M. and Pasca, C. M. (2010), Georges-Louis Leclerc de Buffon's 'Essays on moral arithmetic', *in* A. Sadrieh and A. Ockenfels, eds, 'The Selten School of Behavioral Economics: A Collection of Essays in Honor of Reinhard Selten', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 245–282.

Hong, F. (n.d.), 'What do you know in St. Petersburg? An exploration of "knowledge-first" decision theory'. Unpublished manuscript.

Ichikawa, J. J. and Steup, M. (2018), The Analysis of Knowledge, *in* E. N. Zalta, ed., 'The Stanford Encyclopedia of Philosophy', Summer 2018 edn, Metaphysics Research Lab, Stanford University.

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.

Kreps, D. M. (1988), *Notes on the Theory of Choice*, Westview Press, Boulder.

Lehmann, E. L. (1955), 'Ordered families of distributions', *The Annals of Mathematical Statistics* **26**(3), 399–419.

Lewis, D. (1999), Why conditionalize?, *in* 'Papers in Metaphysics and Epistemology', Vol. 2 of *Cambridge Studies in Philosophy*, Cambridge University Press, Cambridge, pp. 403–407.

Lundgren, B. and Stefánsson, H. O. (2020), 'Against the *De Minimis* principle', *Risk Analysis* **40**(5), 908–914.

MacAskill, W. (2019), 'Longtermism', Effective Altruism Forum.
URL: *https://forum.effectivealtruism.org/posts/qZyshHCNkjs3TvSem/longtermism*

Mann, H. B. and Whitney, D. R. (1947), 'On a test of whether one of two random variables is stochastically larger than the other', *The Annals of Mathematical Statistics* **18**(1), 50–60.

McMahan, J. (1981), 'Problems of population theory', *Ethics* **92**(1), 96–127.

Menger, K. (1934), 'Das unsicherheitsmoment in der wertlehre: Betrachtungen im anschliß an das sogenannte petersburger spiel', *Zeitschrift für Nationalökonomie / Journal of Economics* **5**(4), 459–485.

Menger, K. (1967), The role of uncertainty in economics, *in* M. Shubik, ed., 'Essays in Mathematical Economics: In Honor of Oskar Morgenstern', Princeton University Press, Princeton, pp. 211–231.

Monton, B. (2019), 'How to avoid maximizing expected utility', *Philosophers' Imprint* **19**(18), 1–24.

Nebel, J. M. (2019), 'An intrapersonal addition paradox', *Ethics* **129**(1), 309–343.

Nover, H. and Hájek, A. (2004), 'Vexing expectations', *Mind* **113**(450), 237–249.

Parfit, D. (1984), *Reasons and Persons*, Clarendon Press, Oxford.

Pascal, B. (1958), *Pascal's Pensées*, E. P. Dutton & Co., New York.
URL: *https://www.gutenberg.org/files/18269/18269-0.txt*

Peterson, M. (2002), 'What is a *de minimis* risk?', *Risk Management* **4**(2), 47–55.

Pulskamp, R. J. (2013), 'Correspondence of Nicolas Bernoulli concerning the St. Petersburg Game'. Unpublished manuscript. Accessed through: https://web.archive.org/.
URL: *http://cerebro.xu.edu/math/Sources/NBernoulli/correspondence_petersburg_game.pdf*

Quirk, J. P. and Saposnik, R. (1962), 'Admissibility and measurable utility functions', *The Review of Economic Studies* **29**(2), 140–146.

Russell, J. S. (2021), 'On two arguments for fanaticism'. Global Priorities Institute Working Paper No. 17–2021.

URL: *https://globalprioritiesinstitute.org/on-two-arguments-for-fanaticism-jeff-sanford-russell-university-of-southern-california/*

Russell, J. S. and Isaacs, Y. (2021), 'Infinite prospects', *Philosophy and Phenomenological Research* **103**(1), 178–198.

Samuelson, P. A. (1977), 'St. petersburg paradoxes: Defanged, dissected, and historically described', *Journal of Economic Literature* **15**(1), 24–55.

Savage, L. J. (1972), *The Foundations of Statistics*, 2 edn, Dover Publications, New York.

Smith, M. (2014*a*), 'What else justification could be', *Noûs* **44**(1), 10–31.

Smith, N. J. J. (2014*b*), 'Is evaluative compositionality a requirement of rationality?', *Mind* **123**(490), 457–502.

Smith, N. J. J. (2016), 'Infinite decisions and rationally negligible probabilities', *Mind* **125**(500), 1199–1212.

Tarsney, C. (2020), 'Exceeding expectations: Stochastic dominance as a general decision theory'. Global Priorities Institute Working Paper No. 3–2020.
URL: *https://globalprioritiesinstitute.org/christian-tarsney-exceeding-expectations-stochastic-dominance-as-a-general-decision-theory/*

von Neumann, J. and Morgenstern, O. (1947), *Theory of Games and Economic Behavior*, 2 edn, Princeton University Press, Princeton.

Wilkinson, H. (2022), 'In defence of fanaticism', *Ethics* **132**(2), 445–477.

Yudkowsky, E. (2007*a*), '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*

Zynda, L. (2000), 'Representation theorems and realism about degrees of belief', *Philosophy of Science* **67**(1), 45–69.