The St. Petersburg Paradox

Peterson, Martin

1. The History of the St. Petersburg Paradox

The St. Petersburg paradox is named after one of the leading scientific journals of the eighteenth century, *Commentarii Academiae Scientiarum Imperialis Petropolitanae* [*Papers of the Imperial Academy of Sciences in Petersburg*], in which Daniel Bernoulli (1700–1782) published a paper entitled "Specimen Theoriae Novae de Mensura Sortis" ["Exposition of a New Theory on the Measurement of Risk"] in 1738. Daniel Bernoulli had learned about the problem from his cousin Nicolaus I (1687–1759), who proposed an early but unnecessarily complex version of the paradox in a letter to Pierre Rémond de Montmort on 9 September 1713 (for this and related letters see J. Bernoulli 1975). Nicolaus asked de Montmort to imagine an example in which an ordinary dice is rolled until a 6 comes up:

> [W]hat is the expectation of *B* ... if *A* promises to *B* to give him some coins in this progression 1, 2, 4, 8, 16 etc. or 1, 3, 9, 27 etc. or 1, 4, 9, 16, 25 etc. or 1, 8, 27, 64 instead of 1, 2, 3, 4, 5 etc. as beforehand. Although for the most part these problems are not difficult, you will find however something most curious. (N. Bernoulli to Montmort, 9 September 1713)

It seems that Montmort did not immediately get Nicolaus' point. Montmort responded that these problems

> have no difficulty, the only concern is to find the sum of the series of which the numerators being in the progression of squares, cubes, etc. the denominators are in geometric progression. (Montmort to N. Bernoulli, 15 November 1713)

However, he never performed any calculations. If he had, he would have discovered that the expected value of the first series (1, 2, 4, 8, 16, etc.) is:

$$\sum_{n=1}^{\infty} \frac{5^{n-1}}{6^n} \cdot 2^{n-1}.$$

For this series it holds that

$$\lim_{n \to \infty} \left| \frac{a_{n+1}}{a_n} \right| > 1,$$

so by applying the ratio test it is easy to verify that the series is divergent. (This test was discovered by d'Alembert in 1768, so it might be unfair to criticize Montmort for not seeing this.) However, the mathematical argument presented by Nicolaus himself was also a bit sketchy and would not impress contemporary mathematicians. The good news is that his conclusion was correct:

> it would follow thence that B must give to *A* an infinite sum and even more than infinity (if it is permitted to speak thus) in order that he be able to make the advantage to give him some coins in this progression 1, 2, 4, 8, 16 etc. (N. Bernoulli to Montmort, 20 February 1714)

The next important contribution to the debate was made by Cramér in 1728. He read about Nicolaus' original problem in a book published by Montmort and proposed a simpler and more elegant formulation in a letter to Nicolaus:

> In order to render the case more simple I will suppose that *A* throw in the air a piece of money, *B* undertakes to give him a coin, if the side of Heads falls on the first toss, 2, if it is only the second, 4, if it is the 3rd toss, 8, if it is the 4th toss, etc. The paradox consists in this that the calculation gives for the equivalent that *A* must give to *B* an infinite sum, which would seem absurd. (Cramér to N. Bernoulli, 21 May 1728)

In the very same letter, Cramér proposed a solution that revolutionized the emerging field of decision theory. Cramér pointed out that it is not the expected *monetary* value that should guide the choices of a rational agent, but rather the "usage" that "men of good sense" can make of money. According to Cramér, twenty million is not worth more than ten million, because ten million is enough for satisfying all desires an agent may reasonably have:

> mathematicians value money in proportion to its quantity, and men of good sense in proportion to the usage that they may make of it. That which renders the mathematical expectation infinite, is the prodigious sum that I am able to receive, if the side of Heads falls only very late, the 100th or 1000th toss. Now this sum, if I reason as a sensible man, is not more for me, does not make more pleasure for me, does not engage me more to accept the game, than if it would be only 10 or 20 million coins. (21 May 1728)

The point made by Cramér in this passage can be generalized. Suppose that the upper boundary of an outcome's value is 2^m. If so, that outcome will be obtained if the coin lands heads on the mth flip. This means that the expected value of all the *infinitely*

many possible outcomes in which the coin is flipped more than m times will be finite: It is 2^m times the probability that this happens, so it cannot exceed 2^m. To this we have to add the aggregated value of the first m possible outcomes, which is obviously finite. Because the sum of any two finite numbers is finite, the expected value of Cramér's version of the St. Petersburg game is finite.

Cramér was aware that it would be controversial to claim that there exists an upper boundary beyond which additional riches do not matter *at all*. However, he pointed out that his solution works even if he the value of money is strictly increasing but the relative increase gets smaller and smaller (21 May 1728):

> If one wishes to suppose that the moral value of goods was as the square root of the mathematical quantities … my moral expectation will be

$$\frac{1}{2} \cdot \sqrt{1} + \frac{1}{4} \cdot \sqrt{2} + \frac{1}{8} \cdot \sqrt{4} + \frac{1}{16} \cdot \sqrt{8} \ldots$$

This is the first clear statement of what contemporary decision theorists and economists refer to as decreasing marginal utility: The additional utility of more money is never zero, but the richer you are, the less you gain by increasing your wealth further. Cramér correctly calculated the expected utility ("moral value") of the St. Petersburg game to be about 2.9 units for an agent whose utility of money is given by the root function.

Daniel Bernoulli proposed a very similar idea in his famous 1738 article mentioned at the beginning of this section. Daniel argued that the agent's utility of wealth equals the logarithm of the monetary amount, which entails that improbable but large monetary prizes will contribute less to the expected utility of the game than more probable but smaller monetary amounts. As his article was about to be published, Daniel's brother Nicolaus mentioned to him that Cramér had proposed a very similar idea in 1728 (in the letter quoted above). In the final version of the text, Daniel openly acknowledged this:

> Indeed I have found [Cramér's] theory so similar to mine that it seems miraculous that we independently reached such close agreement on this sort of subject. (Daniel Bernoulli 1738 [1954: 33])

2. The Modern St. Petersburg Paradox

Cramér's remark about the agent's decreasing marginal utility of money solves the original version of the St. Petersburg paradox. However, modern decision theorists agree that this solution is too narrow. The paradox can be restored by increasing the values of the outcomes up to the point at which the agent is fully compensated for her decreasing marginal utility of money (see Menger 1934 [1979]). The version of the St.

Petersburg paradox discussed in the modern literature can thus be formulated as follows:

> A fair coin is flipped until it comes up heads. At that point the player wins a prize worth 2^n *units of utility* on the player's personal utility scale, where *n* is the number of times the coin was flipped.

Note that the expected utility of this gamble is infinite even if the agent's marginal utility of money is decreasing. We can leave it open exactly what the prizes consists of. It need not be money.

It is worth stressing that none of the prizes in the St. Petersburg game have infinite value. No matter how many times the coin is flipped, the player will *always win some finite amount of utility*. The *expected* utility of the St. Petersburg game is not finite, but the *actual* outcome will always be finite. It would thus be a mistake to dismiss the paradox by arguing that no actual prizes can have infinite utility. No actual infinities are required for constructing the paradox, only potential ones. (For a discussion of the distinction between actual and potential infinities, see Linnebo and Shapiro 2019.) In discussions of the St. Petersburg paradox it is often helpful to interpret the term "infinite utility" as "not finite" but leave it to philosophers of mathematics to determine whether it *is* or merely *approaches* infinity.

Some authors have discussed exactly what is problematic with the claim that the expected utility of the modified St. Petersburg game is infinite (read: not finite). Is it merely the fact that the fair price of the wager is "too high", or is there something else that prompts the worry? James M. Joyce notes that

> a wager of infinite utility will be strictly preferred to *any* of its payoffs since the latter are all finite. This is absurd given that we are confining our attention to bettors who value wagers only as means to the end of increasing their fortune. (Joyce 1999: 37)

Joyce's point seems to be that an agent who pays the fair price of the wager will know *for sure* that she will actually be worse off after she has paid the fee. However, this seems to presuppose that actual infinities do exist. If only potential infinities exist, then the player cannot "pay" an infinite fee for playing the game. If so, we could perhaps interpret Joyce as reminding us that no matter what finite amount the player actually wins, the expected utility will always be higher, meaning that it would have been rational to pay even more. Russell and Isaacs (2021:179) offer a slightly different analysis. Their point is that "however much the St. Petersburg gamble is worth, no particular *outcome* could be worth exactly that much". This is because the St. Petersburg gamble is worth more than any finite outcome, but less than something worth infinitely much.

> Is the St. Petersburg gamble perhaps worth something like an infinite

amount of money? No. The St. Petersburg gamble is sure to pay only a finite amount of money. Suppose there is something which is worth *more* than each finite amount of money–such as an infinite amount of money (whatever that might come to), or a priceless artwork, or true love. If [the agent] has something like that, then the prospect of keeping it (with certainty) will *dominate* giving it up in exchange for the St. Petersburg gamble; thus the St. Petersburg gamble is not worth so much. Of course, nothing is worth *more* than each finite amount of money, yet not worth more than *every* finite amount of money. So the conclusion of [the agent's] reasoning is that nothing she could bid–monetary or otherwise– would be the right price. (Russell and Isaacs 2021:179)

Decisions theorists wish to clarify a means-ends notion of rationality, according to which it is rational to do whatever is the best means to one's end. The player thus knows that paying more than what one actually wins *cannot* be the best means to the end of maximizing utility. But always being forced to pay too little is also problematic, because then the seller would "pay" too much (that is, receive too little). So at least one agent will be irrational and pay too much unless we can establish a fair price of the gamble. This observation enables us to strengthen the original "paradox" (in which no formal contradiction is derived) into a stronger version consisting of three incompatible claims:

1. The amount of utility it is rational to pay for playing (or selling the right to play) the St. Petersburg game is higher than every finite amount of utility.
2. The buyer knows that the actual amount of utility he or she will actually receive is finite.
3. It is not rational to knowingly pay more for something than one will receive.

Many discussions of the St. Petersburg paradox have focused on (1). As we will see in the next couple of sections, many scholars argue that the value of the St. Petersburg game is, for one reason or another, finite. A rare exception is Hájek and Nover. They offer the following argument for accepting (1):

The St Petersburg game can be regarded as the limit of a sequence of truncated St Petersburg games, with successively higher finite truncation points—for example, the game is called off if heads is not reached by the tenth toss; by the eleventh toss; by the twelveth toss;…. If we accept dominance reasoning, these successive truncations can guide our assessment of the St Petersburg game's value: it is bounded below by each of their values, these bounds monotonically increasing. Thus we have a principled reason for accepting that it is worth paying any finite amount to play the St Petersburg game. (Hájek and Nover 2006: 706)

Although they do not explicitly say so, Hájek and Nover would probably reject (3). The least controversial claim is perhaps (2). It is, of course, *logically possible* that the coin

keeps landing tails *every* time it is flipped, even though an infinite sequence of tails has probability 0. (For a discussion of this possibility, see Williamson 2007.) Some events that have probability 0 do actually occur, and in uncountable probability spaces it is impossible that all outcomes have a probability greater than 0. Even so, if the coin keeps landing tails *every* time it is flipped, the agent wins 0 units of utility. So [(2)](#) would still hold true.

3. Unrealistic Assumptions?

Some authors claim that the St. Petersburg game should be dismissed because it rests on assumptions that can never be fulfilled. For instance, Jeffrey (1983: 154) argues that "anyone who offers to let the agent play the St. Petersburg gamble is a liar, for he is pretending to have an indefinitely large bank". Similar objections were raised in the eighteenth century by Buffon and Fontaine (see Dutka 1988).

However, it is not clear why Jeffrey's point about real-world constraints would be relevant. What is wrong with evaluating a highly idealized game we have little reason to believe we will ever get to play? Hájek and Smithson (2012) point out that the St Petersburg paradox is *contagious* in the following sense: As long as you assign some nonzero probability to the hypothesis that the bank's promise is credible, the expected utility will be infinite no matter how low your credence in the hypothesis is. Any nonzero probability times infinity equals infinity, so any option in which you get to play the St. Petersburg game with a nonzero probability has infinite expected utility.

It is also worth keeping in mind that the St. Petersburg game may not be as unrealistic as Jeffrey claims. The fact that the bank does not have an indefinite amount of money (or other assets) available *before the coin is flipped* should not be a problem. All that matters is that the bank can make *a credible promise* to the player that the correct amount will be made available within a reasonable period of time after the flipping has been completed. How much money the bank has in the vault when the player plays the game is irrelevant. This is important because, as noted in section 2, the amount the player actually wins will always be finite. We can thus imagine that the game works as follows: We first flip the coin, and once we know what finite amount the bank owes the player, the CEO will see to it that the bank raises enough money.

If this does not convince the player, we can imagine that the central bank issues a blank check in which the player gets to fill in the correct amount once the coin has been flipped. Because the check is issued by the central bank it cannot bounce. New money is automatically created as checks issued by the central bank are introduced in the economy. Jeffrey dismisses this version of the St. Petersburg game with the following argument:

> [Imagine that] Treasury department delivers to the winner a crisp new billion billion dollar bill. Due to the resulting inflation, the marginal

> desirabilities of such high payoffs would presumably be low enough to
> make the prospect of playing the game have finite expected [utility].
> (Jeffrey 1983: 155)

Jeffrey is probably right that "a crisp new billion billion dollar bill" would trigger some inflation, but this seems to be something we could take into account as we construct the game. All that matters is that the utilities in the payoff scheme are linear.

Readers who feel unconvinced by this argument may wish to imagine a version of the St. Petersburg game in which the player is hooked up to Nozick's Experience Machine (see section 2.3 in the entry on hedonism). By construction, this machine can produce any pleasurable experience the agent wishes. So once the coin has been flipped n times, the Experience Machine will generate a pleasurable experience worth 2^n units of utility on the player's personal utility scale. Aumann (1977) notes without explicitly mention the Experience Machine that:

> The payoffs need not be expressible in terms of a fixed finite number of
> commodities, or in terms of commodities at all […] the lottery ticket […]
> might be some kind of open-ended activity -- one that could lead to
> sensations that he has not heretofore experienced. Examples might be
> religious, aesthetic, or emotional experiences, like entering a monastery,
> climbing a mountain, or engaging in research with possibly spectacular
> results. (Aumann 1977: 444)

A possible example of the type of experience that Aumann has in mind could be the number of days spent in Heaven. It is not clear why time spent in Heaven must have diminishing marginal utility.

Another type of practical worry concerns the temporal dimension of the St. Petersburg game. Brito (1975) claims that the coin flipping may simply take too long time. If each flip takes n seconds, we must make sure it would be *possible* to flip it sufficiently many times before the player dies. Obviously, if there exists an upper limit to how many times the coin can be flipped the expected utility would be finite too.

A straightforward response to this worry is to imagine that the flipping took place yesterday and was recorded on video. The first flip occurred at 11 p.m. sharp, the second flip $\frac{60}{2}$ minutes later, the third $\frac{60}{4}$ minutes after the second, and so on. The video has not yet been made available to anyone, but as soon as the player has paid the fee for playing the game the video will be placed in the public domain. Note that the coin could in principle have been flipped infinitely many times within a single hour. (This is an example of a "supertask"; see the entry on supertasks.)

It is true that this random experiment requires the coin to be flipped faster and faster. At some point we would have to spin the coin faster than the speed of light. This is not

logically impossible although this assumption violates a contingent law of nature. If you find this problematic, we can instead imagine that someone throws a dart on the real line between 0 and 1. The probability that the dart hits the first half of the interval, $\left[0, \frac{1}{2}\right)$, is $\frac{1}{2}$. And the probability that the dart hits the next quarter, $\left[\frac{1}{2}, \frac{3}{4}\right)$, is $\frac{1}{4}$, and so on. If "coin flips" are generated in this manner the random experiment will be over in no time at all. To steer clear of the worry that no real-world dart is infinitely sharp we can define the point at which the dart hits the real line as follows: Let a be the area of the dart. The point at which the dart hits the interval [0,1] is defined such that half of the area of a is to the right of some vertical line through a and the other half to the left the vertical line. The point at which the vertical line crosses the interval [0,1] is the outcome of the random experiment.

In the contemporary literature on the St. Petersburg paradox practical worries are often ignored, either because it is possible to imagine scenarios in which they do not arise, or because highly idealized decision problems with unbounded utilities and infinite state spaces are deemed to be interesting in their own right.

4. A Bounded Utility Function?

Arrow (1970: 92) suggests that the utility function of a rational agent should be "taken to be a bounded function…. since such an assumption is needed to avoid [the St. Petersburg] paradox". Basset (1987) makes a similar point; see also Samuelson (1977) and McClennen (1994).

Arrow's point is that utilities must be bounded to avoid the St. Petersburg paradox and that traditional axiomatic accounts of the expected utility principle guarantee this to be the case. The well-known axiomatizations proposed by Ramsey (1926), von Neumann and Morgenstern (1947), and Savage (1954) do, for instance, all entail that the decision maker's utility function is bounded. (See [section 2.3 in the entry on decision theory](#) for an overview of von Neumann and Morgenstern's axiomatization.)

If the utility function is bounded, then the expected utility of the St. Petersburg game will of course be finite. But why do the axioms of expected utility theory guarantee that the utility function is bounded? The crucial assumption is that rationally permissible preferences over lotteries are *continuous*. To explain the significance of this axiom it is helpful to introduce some symbols. Let $\{pA, (1 - p)B\}$ be the lottery that results in A with probability p and B with probability $1 - p$. The expression $A \preceq B$ means that the agent considers B to be at least as good as A, i.e., weakly prefers B to A. Moreover, $A \sim B$ means that A and B are equi-preferred, and $A \prec B$ means that B is preferred to A. Consider:

- *The Continuity Axiom:* Suppose $A \preceq B \preceq C$. Then there is a probability $p \in [0, 1]$

such that $\{pA, (1 - p)C\} \sim B$.

To explain why this axiom entails that no object can have infinite value, suppose for *reductio* that *A* is a prize check worth \$1, *B* is a check worth \$2, and *C* is a prize to which the agent assigns infinite utility. The decision maker's preference is $A \prec B \prec C$, but there is no probability *p* such that $\{pA, (1 - p)C \sim B$. Whenever *p* is nonzero the decision maker will strictly prefer $\{pA, (1 - p)C\}$ to *B*, and if *p* is 0 the decision maker will strictly prefer *B*. So because no object (lottery or outcome) can have infinite value, and a utility function is defined by the utilities it assigns to those objects (lotteries or outcomes), the utility function has to be bounded.

Does this solve the St. Petersburg paradox? The answer depends on whether we think a rational agent offered to play the St. Petersburg game has any reason to accept the continuity axiom. A possible view is that anyone who is offered to play the St. Petersburg game has reason to reject the continuity axiom. Because the St. Petersburg game has infinite utility, the agent has no reason to evaluate lotteries in the manner stipulated by this axiom. As explained in Section 3, we can imagine unboundedly valuable payoffs.

Some might object that the continuity axiom, as well as the other axioms proposed by von Neumann and Morgenstern (and Ramsey and Savage), are essential for *defining* utility in a mathematically precise manner. It would therefore be *meaningless* to talk about utility if we reject the continuity axiom. This axiom is part of what it means to say that something has a higher utility than something else. A good response could be to develop a theory of utility in which preferences over lotteries are not used for defining the meaning of the concept; see Luce (1959) for an early example of such a theory. Another response could be to develop a theory of utility in which the continuity axiom is explicitly rejected; see Skala (1975).

5. Ignore Small Probabilities?

Buffon argued in 1777 that a rational decision maker should disregard the possibility of winning lots of money in the St. Petersburg game because the probability of doing so is very low. According to Buffon, some sufficiently improbable outcomes are "morally impossible" and should therefore be ignored. From a technical point of view, this solution is very simple: The St. Petersburg paradox arises because the decision maker is willing to aggregate infinitely many extremely valuable but highly improbable outcomes, so if we restrict the set of "possible" outcomes by excluding sufficiently improbable ones the expected utility will, of course, be finite.

But *why* should small probabilities be ignored? And how do we draw the line between small probabilities that are beyond concern and others that are not? Dutka summarizes Buffon's lengthy answer as follows:

To arrive at a suitable threshold value, [Buffon] notes that a fifty-six year old man, believing his health to be good, would disregard the probability that he would die within twenty-four hours, although mortality tables indicate that the odds against his dying in this period are only 10189 to 1. Buffon thus takes a probability of 1/10,000 or less for an event as a probability which may be disregarded. (Dutka 1988: 33)

Is this a convincing argument? According to Buffon, we *ought* to ignore some small probabilities because people like him (56-year-old males) *do in fact* ignore them. Buffon can thus be accused of attempting to derive an "ought" from an "is". To avoid Hume's no-ought-from-an-is objection, Buffon would have to add a premise to the effect that people's everyday reactions to risk are always rational. But why should we accept such a premise?

Another objection is that if we ignore small probabilities, then we will sometimes have to ignore *all* possible outcomes of an event. Consider the following example: A regular deck of cards has 52 cards, so it can be arranged in exactly 52! different ways. The probability of any given arrangement is thus about 1 in $8 \cdot 10^{67}$. This is a very small probability. (If one were to add six cards to the deck, then the number of possible orderings would exceed the number of atoms in the known, observable universe.) However, every time we shuffle a deck of cards, we know that exactly one of the possible outcomes will materialize, so why should we ignore *all* such very improbable outcomes?

Nicholas J. J. Smith (2014) defends a modern version of Buffon's solution. He bases his argument on the following principle:

- *Rationally negligible probabilities (RNP):* For any lottery featuring in any decision problem faced by any agent, there is an $\epsilon > 0$ such that the agent need not consider outcomes of that lottery of probability less than ϵ incoming to a fully rational decision. (Smith 2014: 472)

Smith points out that the order of the quantifiers in RNP is crucial. The claim is that for every lottery there exists some probability threshold ϵ below which all probabilities should be ignored, but it would be a mistake to think that one and the same ϵ is applicable to every lottery. This is important because otherwise we could argue that RNP allows us to combine thousands or millions of separate events with a probability of less than ϵ. It would obviously make little sense to ignore, say, half a million one-in-a-million events. By keeping in mind that that the appropriate ϵ may vary from case to case this worry can be dismissed.

Smith also points out that if we ignore probabilities less than ϵ, then we have to increase some other probabilities to ensure that all probabilities sum up to one, as

required by the probability axioms (see). Smith proposes a principle for doing this in a systematic manner.

However, *why* should we accept RNP? What is the *argument* for accepting this controversial principle apart from the fact that it would solve the St. Petersburg paradox? Smith's argument goes as follows:

> Infinite precision cannot be required: rather, in any given context, there must be some finite tolerance—some positive threshold such that ignoring all outcomes whose probabilities lie below this threshold counts as satisfying the norm…. There is a norm of decision theory which says to ignore outcomes whose probability is zero. Because this norm mentions a specific probability value (zero), it is the kind of norm where it makes sense to impose a tolerance: zero plus or minus ϵ (which becomes zero plus ϵ, given that probabilities are all between 0 and 1)… the idea behind (RNP) is that in any actual context in which a decision is to be made, one never needs to be infinitely precise in this way—that it never matters. There is (for each decision problem, each lottery therein, and each agent) some threshold such that the agent would not be irrational if she simply ignored outcomes whose probabilities lie below that threshold. (Smith 2014: 472–474)

Suppose we accept the claim that infinite precision is not *required* in decision theory. This would entail, per Smith's argument, that it is *rationally permissible* to ignore probabilities smaller than ϵ. However, to ensure that the decision maker never pays a fortune for playing the St. Petersburg game it seems that Smith would have to defend the stronger claim that decision makers are *rationally required* to ignore small probabilities, i.e., that it is not permissible to not ignore them. Decision makers who find themselves in agreement with Smith's view run a risk of paying a very large amount for playing the St. Petersburg game without doing anything deemed to be irrational by RNP. This point is important because it is arguably more difficult to show that decision makers are *rationally required* to avoid "infinite precision" in decisions in which this is an attainable and fully realistic goal, such as the St. Petersburg game. For a critique of RNP and a discussion of some related issues, see Hájek (2014).

Another objection to RNP has been proposed by Yoaav Isaacs (2016). He shows that RNP together with an additional principle endorsed by Smith (Weak Consistency) entail that the decision maker will sometimes take arbitrarily much risk for arbitrarily little reward.

Lara Buchak (2013) proposes what is arguably a more elegant version of this solution. Her suggestion is that we should assign exponentially *less weight* to small probabilities as we calculate an option's value. A possible weighting function *r*

discussed by Buchak is $r(p) = p^2$. Her proposal is, thus, that if the probability is $\frac{1}{8}$ that you win \$8 in addition to what you already have, and your utility of money increases linearly, then instead of multiplying your gain in utility by $\frac{1}{8}$, you should multiply it by $(\frac{1}{8})^2 = \frac{1}{64}$. Moreover, if the probability is $\frac{1}{16}$ that you win \$16 in addition to what you already have, you should multiply your gain by $\frac{1}{256}$, and so on. This means that small probabilities contribute very little to the *risk-weighted* expected utility.

Buchak's proposal vaguely resembles the familiar idea that our marginal utility of money is decreasing. As stressed by Cramér and Daniel Bernoulli, more money is always better than less, but the utility gained from each extra dollar is decreasing. According to Buchak, the weight we should assign to an outcome's probability is also nonlinear: Small probabilities matter less the smaller they are, and their relative importance decrease exponentially:

> The intuition behind the diminishing marginal utility analysis of risk aversion was that adding money to an outcome is of less value the more money the outcome already contains. The intuition behind the present analysis of risk aversion is that adding probability to an outcome is of more value the more likely that outcome already is to obtain. (Buchak 2014: 1099.)

Buchak notes that this move does not by itself solve the St. Petersburg paradox. For reasons that are similar to those Menger (1934 [1979]) mentions in his comment on Bernoulli's solution, the paradox can be reintroduced by adjusting the outcomes such that the sum increases linearly (for details, see Buchak 2013: 73–74). Buchak is, for this reason, also committed to RNP, i.e., the controversial assumption that there will be some probability so small that it does not make any difference to the overall value of the gamble.

Another worry is that because Buchak rejects the principle of maximizing expected utility and replaces it with the principle of risk-weighted maximizing expected utility, many of the stock objections decision theorists have raised against violations of the expected utility principle can be raised against her principle as well. For instance, if you accept the principle of risk-weighted maximizing expected utility, you have to reject the independence axiom. This entails that you can be exploited in some cleverly designed pragmatic argument. See Briggs (2015) for a discussion of some objections to Buchak's theory.

6. Relative Expected Utility Theory

In the Petrograd game introduced by Colyvan (2008) the player wins \$1 more than in the St. Petersburg game regardless of how many times the coin is flipped. So instead

of winning 2 utility units if the coin lands heads on the first toss, the player wins 3; and so on. See Table 1.

Table 1

Probability	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	...
St. Petersburg	2	4	8	...
Petrograd	2 + 1	4 + 1	8 + 1	...

It seems obvious that the Petrograd game is worth more than the St. Petersburg game. However, it is not easy to explain why. Both games have infinite expected utility, so the expected utility principle gives the wrong answer. It is not true that the Petrograd game is worth more than the St. Petersburg game because its expected utility is higher; the two games have exactly the same expected utility. This shows that the expected utility principle is not universally applicable to all risky choices, which is an interesting observation in its own right.

Is the Petrograd game worth more than the St. Petersburg game because the outcomes of the Petrograd game *dominate* those of the St. Petersburg game? In this context, dominance means that the player will always win $1 more regardless of which state of the world turns out to be the true state, that is, regardless of how many times the coin is flipped. The problem is that it is easy to imagine versions of the Petrograd game to which the dominance principle would not be applicable. Imagine, for instance, a version of the Petrograd game that is exactly like the one in Table 1 except that for some very improbable outcome (say, if the coin lands heads for the first time on the 100th flip) the player wins 1 unit *less* than in the St. Petersburg game. This game, the Petrogradskij game, does not dominate the St. Petersburg game. However, since it is almost certain that the player will be better off by playing the Petrogradskij game a plausible decision theory should be able to explain why the Petrogradskij game is worth more than the St. Petersburg game.

Colyvan claims that we can solve this puzzle by introducing a new version of expected utility theory called Relative Expected Utility Theory (REUT). According to REUT we should calculate the difference in expected utility between the two options for each possible outcome. Formally, the relative expected utility (REU) of act A_k over A_l is

$$REU(A_k, A_l) = \sum_{i=1}^{n} p_i(u_{ki} - u_{li}).$$

According to Colyvan, it is rational to choose A_k over A_l if and only if $REU(A_k, A_l) > 0$.

Colyvan's REUT neatly explains why the Petrograd game is worth more than the St. Petersburg game because the relative expected utility is 1. REUT also explains why the

Petrogradskij game is worth more than the St. Petersburg game: the difference in expected utility is $1 - (\frac{1}{2})^{100}$ which is > 0.

However, Peterson (2013) notes that REUT cannot explain why the *Leningradskij* game is worth more than the *Leningrad* game (see Table 2). The Leningradskij game is the version of the Petrograd game in which the player in addition to receiving a finite number of units of utility also gets to play the St. Petersburg game (SP) if the coin lands heads up in the second round. In the Leningrad game the player gets to play the St. Petersburg game (SP) if the coin lands heads up in the third round.

Table 2

Probability	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$ ...
Leningrad	2	4	8 + SP	16 ...
Leningradskij	2	4 + SP	8	16 ...

It is obvious that the Leningradskij game is worth more than the Leningrad game because the probability that the player gets to play SP as a bonus (which has infinite expected utility) is higher. However, REUT cannot explain why. The difference in expected utility for the state that occurs with probability $\frac{1}{4}$ in Table 2 is $-\infty$ and it is $+\infty$ for the state that occurs with probability $\frac{1}{8}$. Therefore, because $p \cdot \infty = \infty$ for all positive probabilities p, and "$\infty - \infty$" is undefined in standard analysis, REUT cannot be applied to these games.

Bartha (2007) and (2016) proposes a more complex version of relative expected utility theory. In Bartha's theory, the utility of an outcome x is compared to the utility of some alternative outcome y and a basepoint z, which can be chosen arbitrarily as long as x and y are at least as preferred as z. The relative utility of x vis-à-vis y and the base-point z is then defined as the ratio between $u(x) - u(z)$ and the denominator $u(y) - u(z)$; the latter is a "measuring stick" to which $u(x) - u(z)$ is compared. So if $u(x) = 10$, $u(y) = 20$ and $u(z) = 0$, then the relative utility of x vis-à-vis y and the base-point z is $U(x, y; z) = 0.5$.

Bartha's suggestion is to ask the agent to compare the St. Petersburg game to a lottery between two other games. If, for instance, Petrograd⁺ is the game in which the player always wins 2 units more than in the St. Petersburg game regardless of how many times the coin is tossed, then the player could compare the Petrograd game to a lottery between Petrograd⁺ and the St. Petersburg game. By determining for what probabilities p a lottery in which one plays Petrograd⁺ with probability p and the St. Petersburg game with probability $1 - p$ is better than playing the Petrograd game for sure one can establish a measure of the relative value of Petrograd as compared to

Petrograd[+] or St. Petersburg. (For details, see Sect. 5 in Bartha 2016. See also Colyvan and Hájek's 2016 discussion of Bartha's theory.)

An odd feature of Bartha's theory is that two lotteries can have the *same* relative utility even if one is *strictly preferred* to the other; see Bartha (2011: 34–35). This indicates that the relative utilities assigned to lotteries in Bartha's theory are not always choice-guiding.

Let us also mention another, quite simple variation of the original St. Petersburg game, which is played as follows (see Peterson 2015: 87): A manipulated coin lands heads up with probability 0.4 and the player wins a prize worth 2^n units of utility, where n is the number of times the coin was tossed. This game, the Moscow game, is more likely to yield a long sequence of flips and is therefore worth more than the St. Petersburg game, but the expected utility of both games is the same, because both games have infinite expected utility. It might be tempting to say that the Moscow game is more attractive because the Moscow game *stochastically dominates* the St. Petersburg game. (That one game stochastically dominates another game means that for every possible outcome, the first game has at least as high a probability of yielding a prize worth at least u units of utility as the second game; and for some u, the first game yields u with a higher probability than the second.) However, the stochastic dominance principle is inapplicable to games in which there is a small risk that the player wins a prize worth slightly less than in the other game. We can, for instance, imagine that if the coin lands heads on the 100th flip the Moscow game pays one unit less than the St. Petersburg game; in this scenario neither game stochastically dominates the other. Despite this, it still seems reasonable to insist that the game that is almost certain to yield a better outcome (in the sense explained above) is worth more. The challenge is to explain why in a robust and non-arbitrary way.

7. The Pasadena Game

The Pasadena paradox introduced by Nover and Hájek (2004) is inspired by the St. Petersburg game, but the pay-off schedule is different. As usual, a fair coin is flipped n times until it comes up heads for the first time. If n is odd the player wins $(2^n) / n$ units of utility; however, if n is even the player *has to pay* $(2^n) / n$ units. How much should one be willing to pay for playing this game?

If we sum up the terms in the temporal order in which the outcomes occur and calculate expected utility in the usual manner we find that the Pasadena game is worth:

$$\frac{1}{2} \cdot \frac{2}{1} - \frac{1}{4} \cdot \frac{4}{2} + \frac{1}{8} \cdot \frac{8}{3} - \frac{1}{16} \cdot \frac{16}{4} + \frac{1}{32} \cdot \frac{16}{5} - \dots$$

$$= 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \dots$$

$$= \sum_{n} \frac{(-1)^{n-1}}{n}$$

This infinite sum converges to *ln* 2 (about 0.69 units of utility). However, Nover and Hájek point out that we would obtain a very different result if we were to rearrange the order in which the very same numbers are summed up. Here is one of many possible examples of this mathematical fact:

$$1 - \frac{1}{2} - \frac{1}{4} + \frac{1}{3} - \frac{1}{6} - \frac{1}{8} + \frac{1}{5} - \frac{1}{10} - \frac{1}{12} + \frac{1}{7} - \frac{1}{14} - \frac{1}{16} \dots$$

$$= \frac{1}{2}(\ln 2).$$

This is, of course, not news to mathematicians. The infinite sum produced by the Pasadena game is known as the *alternating harmonic series*, which is a *conditionally convergent* series. (A series a_n is conditionally convergent if $\sum_{j=1}^{\infty} a_n$ converges but $\sum_{j=1}^{\infty} |a_n|$ diverges.) Because of a theorem known as the Riemann rearrangement theorem, we know that if an infinite series is conditionally convergent, then its terms can always be rearranged such the sum converges to any finite number, or to $+\infty$ or to $-\infty$.

Nover and Hájek's point is that it seems *arbitrary* to sum up the terms in the Pasadena game in the temporal order produced by the coin flips. To see why, it is helpful to imagine a slightly modified version of the game. In their original paper, Nover and Hájek ask us to imagine that:

> We toss a fair coin until it lands heads for the first time. We have written on consecutive cards your pay-off for each possible outcome. The cards read as follows: (Top card) If the first =heads is on toss #1, we pay you $2. [...] By accident, we drop the cards, and after picking them up and stacking them on the table, we find that they have been rearranged. No matter, you say—obviously the game has not changed, since the pay-off schedule remains the same. The game, after all, is correctly and completely specified by the conditionals written on the cards, and we have merely changed the order in which the conditions are presented. (Nover and Hájek 2004: 237–239)

Under the circumstances described here, we seem to have *no* reason to prefer any particular order in which to sum up the terms of the infinite series. So is the expected value of Pasadena game $\ln 2$ or $\frac{1}{2}(\ln 2)$ or $\frac{1}{3}$ or $-\infty$ or 345.68? All these suggestions seem equally arbitrary. Moreover, the same holds true for the Altadena game, in which every payoff is increased by one dollar. The Altadena game is clearly better than then Pasadena game, but advocates of expected utility theory seem unable to explain why.

The literature on the Pasadena game is extensive. See, e.g., Hájek and Nover (2006), Fine (2008), Smith (2014), and Bartha (2016). A particularly influential solution is due to Easwaran (2008). He introduces a distinction between a strong and a weak version of the expected utility principle, inspired by the well-known distinction between the strong and weak versions of the law of large numbers. According to the strong law of large numbers, the average utility of a game converges to its expected utility with probability one as the number of iterations goes to infinity. The weak law of large numbers holds that for a sufficiently large set of trials the probability can be made arbitrarily small that that the average utility will not differ from the expected utility by more than some small pre-specified amount. So according to the weak expected utility principle,

> by fixing in advance a high enough number of n plays, the average payoff per play can be almost guaranteed to be arbitrarily close to $\ln 2$,

while the strong version of the principle entails that

> if one player keeps getting to decide whether to play again or quit, then she can almost certainly guarantee as much profit as she wants, regardless of the (constant) price per play. (Easwaran 2008: 635)

Easwaran's view is that the weak expected utility principle should guide the agent's choice and that the fair price to pay is $\ln 2$.

However, Easwaran's solution cannot be generalized to other games with slightly different payoff schemes. Bartha (2016: 805) describes a version of the Pasadena game that has no expected value. In this game, the Arroyo game, the player wins $-1^{n+1}(n+1)$ with probability $p_n = \frac{1}{7}(n+1)$. If we calculate the expected utility in the order in which the outcomes are produced, we get the same result as for the Pasadena game: $1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4}\cdots$ For reasons explained (and proved) by Bartha, the Arroyo game has no weak expected utility.

It is also worth keeping in mind that Pasadena-like scenarios can arise in non-probabilistic contexts (see Peterson 2013). Imagine, for instance, an infinite population in which the utility of individual number j is $\frac{(-1)^{j-1}}{j}$. What is the total

utility of this population? Or imagine that you are the proud owner of a Jackson Pollock painting. An art dealer tells you the overall *aesthetic* value of the painting is the sum of some of its parts. You number the points in the painting with *arbitrary* numbers 1, 2, 3, … (perhaps by writing down the numbers on cards and then dropping all cards on the floor); the aesthetic value of each point j is $\frac{(-1)^{j-1}}{j}$. What is the total aesthetic value of the painting? These examples are non-probabilistic versions of the Pasadena problem, to which the expected utility principle is inapplicable. There is no uncertainty about any state of nature; the decision maker knows for sure what the world is like. This means that Easwaran's distinction between weak and strong expectations is not applicable.

Although some of these problems may appear to be somewhat esoteric, we cannot dismiss them. All Pasadena-like problems are vulnerable to the same *contagion problem* as the St Petersburg game (see [section 2](#)). Hájek and Smithson offer the following colorful illustration:

> You can choose between pizza and Chinese for dinner. Each option's desirability depends on how you weigh probabilistically various scenarios (burnt pizza, perfectly cooked pizza,… over-spiced Chinese, perfectly spiced Chinese…) and the utilities you accord them. Let us stipulate that neither choice dominates the other, yet it should be utterly straightforward for you to make a choice. *But it is not if the expectations of pizza and Chinese are contaminated by even a miniscule [sic] assignment of credence to the Pasadena game.* If the door is opened to it just a crack, it kicks the door down and swamps all expected utility calculations. You cannot even choose between pizza and Chinese. (Hájek and Smithson 2012: 42, emph. added.)

Colyvan (2006) suggests that we should bite the bullet on the Pasadena game and accept that it has *no* expected utility. The contagion problem shows that if we were to do so, we would have to admit that the principle of maximizing expected utility would be applicable to nearly *no* decisions. Moreover, because the contagion problem is equally applicable to all games discussed in this entry (St. Petersburg, Pasadena, Arroyo, etc.) it seems that all these problems may require a unified solution.

For hundreds of years, decision theorists have agreed that rational agents should maximize expected utility. The discussion has mostly been focused on how to interpret this principle, especially for choices in which the causal structure of the world is unusual. However, until recently no one has seriously questioned that the principle of maximizing expected utility is the right principle to apply. The rich and growing literature on the many puzzles inspired by the St. Petersburg paradox indicate that this might have been a mistake. Perhaps the principle of maximizing expected utility should be replaced by some entirely different principle?