manuscript No.
(will be inserted by the editor)

Menger 1934 revisited

Ole Peters

Abstract Karl Menger's 1934 paper on the St. Petersburg paradox contains mathematical errors that invalidate his conclusion that unbounded utility functions, specifically Bernoulli's logarithmic utility, fail to resolve modified versions of the St. Petersburg paradox.

Keywords Menger · unbounded utility · Bernoulli · St. Petersburg paradox · ergodicity

1 Preview and preliminaries

In Section 2 the St. Petersburg paradox is defined. Section 3 motivates its recent resolution using the concept of ergodicity, including Section 3.1 where the ergodicity argument is related to a discussion in the economics literature. Section 4 contrasts this with D. Bernoulli's traditional 1738 resolution, and Section 5 summarizes Menger's 1934 study, setting out its reception and conceptual significance. Section 6 focuses on Menger's errors, with the central result in Section 6.2 that Menger overlooked a second divergence in the application of Bernoulli's resolution to his modified paradox. Section 7 investigates Menger's game using the ergodicity resolution, and Section 8 is a brief summary of the findings of the present study.

1.1 Text and notation

Page numbers for [Menger(1934)] refer to the English translation by Schoellkopf and Mellon from 1967 [Menger(1967)].

O. Peters
Department of Mathematics and Grantham Institute for Climate Change
Imperial College London
180 Queens Gate
SW7 2AZ
E-mail: ole@santafe.edu

I will adopt the notation used there with the following exceptions. Whereas Menger uses the same symbol D, and sometimes D_n for different functions of n, I use $D(n)$ for the general case, $D_A(n)$ for the payout function used by Bernoulli, $D_B(n)$ for that used by Menger, and $D_C(n)$ for a milder version of Menger's game, see Table 1. I introduce the symbol P for the ticket price, which – significantly – does not appear in Menger's equations.

The printing of [Menger(1967)] contains several typographical errors, especially on p. 217, where the intended e^{2^n} is repeatedly misprinted as e^{2n}. Also on this page, $1/2^n$ is misprinted as $\frac{1}{2}/2^n$. No further relevant discrepancies with the original German text were found.

When this is convenient, monetary units are called dollars, \$, for dimensional consistency, although the choice of unit or currency is irrelevant. I will synonymously use the terms "expectation value" and "ensemble average".

Page numbers for [Bernoulli(1738)] refer to the 1956 English translation by Sommer.

2 The St. Petersburg paradox

N. Bernoulli, in a letter to Montmort in 1713 [Montmort(1713)], introduced lotteries of a certain type. D. Bernoulli, in a study published by the Imperial Academy of Sciences in St. Petersburg, considered the following specific case [Bernoulli(1738)]: a fair coin is tossed until the first heads event occurs. The number of coin tosses necessary to arrive at this event is n, and the payout as a function of n is $\$D_A(n) = \$2^{n-1}$. We will refer to this as *game A*. The expectation value of the payout in *game A* is given by the divergent sum

$$\langle D_A(n)\rangle = \sum_{n=1}^{\infty}\left(\frac{1}{2}\right)^n 2^{n-1} = \frac{1}{2}+\frac{1}{2}+\frac{1}{2}+\cdots. \tag{1}$$

N. Bernoulli mentioned that he found this game "curious" [Montmort(1713)], p. 402. Later researchers found it "paradoxical" that, in general, individuals offered to purchase a ticket in this lottery are not willing to pay very much for it, namely no more than a few dollars, rather than their entire fortunes or indeed all the money they can borrow. The surprise by these early researchers reflects the belief that risky ventures may be judged by

Criterion i:
a gamble is worth taking if the expectation value of the net change of wealth, here $\langle D(n)\rangle - P$, is positive.

This criterion fails in the St. Petersburg paradox, in the sense that there is no finite price P at which it discourages participation. *Criterion i* may be attributed to Huygens, who wrote "if any one should put 3 shillings in one hand without telling me which, and 7 in the other, and give me choice of either of them; I say, it is the same thing as if he should give me 5 shillings..." [Huygens(1657)]. Many resolutions were put forward, as reviewed *e.g.* in Menger(1934) and Samuelson(1977).

3 Resolution based on Ergodicity

The present comment is written in view of the connection between ergodic theory and the paradox. The St. Petersburg paradox rests on the apparent contradicton between a positively infinite expectation value of winnings in a game and real people's unwillingness to pay much to be allowed to participate in the game. Bernoulli (1738), p. 24, pointed out that because of this incongruence, the expectation value of net winnings, *criterion i*, has to be "discarded" as a descriptive or prescriptive behavioral rule. One can now decide what to change about "the expectation value of net winnings": either "the expectation value" or the "net winnings" (or both). Bernoulli, see Sec. 4, chose to replace the net winnings by introducing utility. An alternative resolution, motivated by the development of the field of ergodic theory in the late 19th and throughout the 20th century, replaces the expectation value (or ensemble average) with a time average, without introducing utility. Details of this resolution can be found in [Peters(2011)]. Conceptually, ergodic theory deals with the question whether expectation values, which can be thought of as averages over non-interacting copies of a system (sometimes called parallel universes), are identical to time averages, where the dynamics of a single system are averaged along a time trajectory. It is pointed out in [Peters(2011)] that the system under investigation, a mathematical representation of the dynamics of wealth of an individual, is not ergodic, and that this manifests itself as a difference between the ensemble average and the time average of the growth rate of wealth. Historically, ergodic theory emphasizes the relevance of the time average, which was discovered later.

To compute ensemble averages, only a probability distribution is required, whereas time averages require a dynamic. This implies that an additional assumption enters into the resolution in [Peters(2011)]. This assumption corresponds to the multiplicative nature of wealth accumulation: any wealth gained can itself be employed to generate further wealth, which leads to exponential growth. The assumption is in general use, where the word "general" has the same meaning as in the following statement: in general, banks and governments offer exponentially growing interest payments on savings. This multiplicativity leads to a logarithm entering the time-average growth rate. Different assumptions about the dynamics can be envisioned, and these would lead to different functions appearing in place of the logarithm. Consequences of replacing the logarithm with generalized versions of it, in expressions very similar to (Eq. 2), have been a topic of intense study in recent years in statistical mechanics [Hanel et al(2011)Hanel, Thurner, and Gell-Mann, Gell-Mann and Tsallis(2004)].

The treatment acknowledging the non-ergodicity of the system, assuming multiplicative dynamics, produces

Criterion ii:
a gamble is worth taking if $\bar{g}$, the time-average exponential growth rate of a player's wealth, is positive, where

$$\bar{g} = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \ln\left(\frac{W - P + D(n)}{W}\right), \tag{2}$$

where W denotes the wealth of an individual before a ticket is purchased and before any payout is received. For a proof, see [Peters(2011)].

One method of computing both ensemble and time averages is to write down for a finite number of sample members and a finite sampling time an estimator of the quantity of interest. One then considers two different limits: for the ensemble average the number of samples is taken to infinity, and for the time average time is taken to infinity. If the two limiting processes do not commute, then the ensemble average is different from the time average, and the system is manifestly not ergodic [Peters(2011), Peters(2010)].

The encoding by the logarithm of effects of time in an ensemble average ((Eq. 2) is an ensemble average of the logarithm) can be understood by writing the logarithm as a limit, $\ln(x) = \lim_{q\to\infty} q(x^{1/q} - 1)$. This limit corresponds, in the present context, to the lottery being played repeatedly for an infinitely long time in appropriately rescaled time units, see [Peters(2011)]. Thus, in (Eq. 2), which is an ensemble average of a logarithmic growth rate (the time unit being one lottery game), the time-limit is implicitly taken first. Since the two limits do not commute in this case, this is not an ensemble average but a time average.

There is no mention of repetitions of the game in N. Bernoulli's letter, but one side of the paradox is human behavior, shaped by evolution of individuals living through time, making decisions repeatedly in risky situations. It is reasonable to assume that the intuition behind the human behavior is a result of making repeated decisions and considering repeated games.

There is no doubt that in the St. Petersburg paradox the time average is more relevant than the ensemble average: the ensemble average diverges (does not exist). Nor does the interpretation of its divergence as "very large" yield meaningful results. *Criterion i* must therefore indeed be discarded, as noted by Bernoulli. The time average assuming the simplest dynamics, on the other hand, resolves the paradox both mathematically and behaviorally. Mathematically, there exists a finite price at which the gamble should be rejected. Behaviorally, this finite price approximately reflects people's choices, as is evident from Bernoulli's resolution based on behavioral (albeit unsystematic) observations, Sec. 4, which for *game A* is similar to *criterion ii*.

3.1 The time argument in economics

Following [Kelly Jr.(1956)], a debate regarding the use of time averages began in the economics literature. Kelly had computed time-average exponential growth rates in games of chance and argued that utility was not necessary and "too general to shed any light on the specific problems" he considered, [Kelly Jr.(1956)], p. 918. The significance of time averages thus came very close to being fully recognized in economics but the opportunity was missed, and the connection to ergodic theory was not made. This is noteworthy because from the perspective of the ergodicity debate the burden of proof is on him who uses expectation values. Whereas the conceptual meaning of a time average is clear – resulting from a single system whose properties are averaged over time, the use of an ensemble average has to be justified, for instance by showing (for ergodic systems) that it is identical to the time average. Alternatively, an ensemble can be a good approximation if a large sample of essentially

identical and independent copies of the relevant system really exists, as was the case in Boltzmann's studies [Boltzmann(1871),Cohen(1997)].

While the time argument was known [Kelly Jr.(1956),Breiman(1961),Markowitz(1976)], its fundamental character was not recognized. For instance, it was argued to be of limited relevance because it assumes following a system's dynamics for an infinite amount of time [Samuelson(1979)]. Of course, time averages are practically meaningful only if the real system has time to explore the relevant part of its dynamical range. But the point of view of ergodic theory emphasizes the opposite line of argument. To put it provocatively, the ensemble average assumes an infinite number of parallel universes. While real time scales are not infinite, they can be large, whereas the real sample consists of exactly one system (reality) and cannot be enlarged because that would imply the absurdity of creating other universes, or – more mildly – copies of the system.

Due to undetected errors and inaccuracies in [Menger(1934)] it is commonly stated in the economics literature that Menger proved *criterion ii* to be invalid, see below. Because it was not known that *criterion ii* follows from considerations of ergodicity, and that Menger's conclusions are at odds with a well-developed mathematical field, his study was not subjected to sufficient scrutiny, as will become apparent in Section 6.

4 Bernoulli's 1738 resolution

D. Bernoulli, writing about two centuries before the formulation of the ergodic hypothesis, offered the following behavioral resolution of the paradox: Since people care about their monetary wealth only insofar as it is useful to them, he introduced the utility function $U(W)$ that assigns to any wealth $\$W$ a usefulness. Specifically, Bernoulli considered the function $U_B(W) = \ln(W)$. The quantity, Bernoulli suggested, that people consider when deciding whether to take part in the lottery is a combination of the expected gain in their utility and the loss in utility they suffer when they purchase a ticket. This leads to Bernoulli's

Criterion iii:
a gamble is worth taking if the following quantity is positive [Bernoulli(1738)], pp. 26–27:

$$\left\langle \Delta U_B^+ \right\rangle - \Delta U_B^- = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \ln\left(\frac{W+2^{n-1}}{W}\right) - \ln\left(\frac{W}{W-P}\right). \tag{3}$$

The first terms on either side of the equation represent the expected gain in logarithmic utility, resulting from the payouts of the lottery. This would be the net change in utility if tickets were given away for free. The second terms represent the loss in logarithmic utility suffered at the time of purchase, *i.e.* after the ticket is bought but before any payout from the lottery is received. Notice that Bernoulli did not calculate the expected net change in logarithmic utility, which would be

$$\langle \Delta U_B \rangle = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \ln\left(\frac{W+2^{n-1}-P}{W}\right). \tag{4}$$

This expression, (Eq. 4), is mathematically identical to the time-average growth rate, (Eq. 2). The expected net change in logarithmic utility thus encodes in an ensemble average information about time under multiplicative dynamics.

Bernoulli justifies his *criterion iii* as follows: "in a fair game the disutility to be suffered by losing must be equal to the utility to be derived by winning", [Bernoulli(1738)], p. 27. If this leads us to believe that Bernoulli wanted to define a fair game as one where the expected net change in utility is zero, then we will conclude that he made an error in mathematizing this through (Eq. 3). Bernoulli did not explicitly claim to compute the expected net change in utility. Be it by mistake or deliberately, he stated the belief that people tend to act according to *criterion iii*, (Eq. 3). This criterion states that, irrespective of future gains, one would never give away one's entire fortune for a ticket in any game, as this would correspond to an infinite loss in utility at the time of purchase. This could be interpreted as a feeling of distrust. What if the lottery is a scam? It can lead to nonsensical results if a future gain greater than the player's wealth is certain, see Section 6.1. The fact that we are forced to make choices of interpretation is characteristic of the lack of clarity in the debate from the outset. Any statement that is fundamentally behavioral and seems incorrect to us can be interpreted either as a mistake or as a reflection of a different belief regarding human behavior.

The consensus in the literature on utility theory is that Bernoulli meant to compute the expected net change in utility and made a slight error. I have not been able to find an example – other than [Menger(1934)] – where the slight difference between Bernoulli's expression (Eq. 3) and the net change in logarithmic utility, (Eq. 4) was not implicitly "corrected". Already in [Laplace(1814)], p. 439–442, the expected net change in utility is calculated, and the method is ascribed to Bernoulli(1738), whereas (Eq. 3) is not mentioned. The book by [Todhunter(1865)] follows Laplace, as do modern textbooks, which state in one form or another that utility is an object encoding human preferences in its expectation value, *e.g.* [Chernoff and Moses(1959), Samuelson(1983)].

5 Summary and reception of Menger 1934

Sections 1–5 and 7 of [Menger(1934)] provide a review of earlier treatments of the St. Petersburg paradox, Sections 2 and 10 clarifying comments regarding the nature of the paradox, and Sections 8 and 9 a list of behavioral regularities with which resolutions of the paradox should be consistent. In Section 5 Menger criticized D. Bernoulli's 1738 resolution on the grounds of its "*ad hoc* character", referring to the arbitrariness of the chosen utility function.

Menger's main finding:
Menger's paper is best known for the conclusions in its Sections 5 and 6 that only bounded utility functions may be used if positively divergent expected net changes in utility, irrespective of ticket price, are to be avoided in games similar to the St. Petersburg lottery.

This conclusion will be shown to be incorrect. Apart from calling Bernoulli's solution "*ad hoc*" Menger avoids criticizing Bernoulli conceptually and focuses his

efforts on a mathematical critique. As was shown in Section 4, it is indeed necessary to look carefully at Bernoulli's mathematical analysis, but Menger, instead of discovering what was mathematically questionable in [Bernoulli(1738)], was misled by precisely this questionable detail to his incorrect conclusion in his Section 6 that only bounded utility functions are permissible.

5.1 Motivation for the present study

The requirement of boundedness specifically rules out logarithmic utility. It was shown in Section 3, however, that logarithmic utility is mathematically equivalent to the conceptually wholly different resolution based on ergodic theory. This point of view provides a firm basis on which to erect a scientific formalism that does not depend on psychological characteristics of human beings. It allows, for instance, to define and compute optimal values of leverage for investments [Peters(2010)] and precipitates a new notion of efficiency, which may be called "stochastic market efficiency", based on dynamic stability arguments. This type of efficiency has significant predictive power and has been corroborated empirically [Peters and Adamou(2011)]. In this new context, [Menger(1934)] raises the question whether ergodic theory has to be revised, Menger's argument is invalid or, of course, [Peters(2010), Peters(2011), Peters and Adamou(2011)] are incorrect. Close inspection shows that the slight inaccuracy in D. Bernoulli's 1738 computation of an expectation value propagated to the extreme context of Menger's work, where it turns into a significant error and indeed invalidates Menger's argument for the boundedness of utility functions. Menger followed Bernoulli too closely but not carefully enough, copying the central computation unquestioningly but incompletely, whereas other researchers implictly corrected Bernoulli *e.g.* [Laplace(1814)], p. 440, [Todhunter(1865)], p. 221.

5.2 Reception of Menger 1934

Menger's detailed study [Menger(1934)] is widely cited and considered an important milestone in the development of utility theory [Markowitz(1976), Samuelson(1977), Arrow(2009)].

It is interesting that the weakness of Menger's argument remained undiscovered for at least 77 years (the argument was prepared in 1923, presented in 1927 and published in 1934 [Menger(1934)]). Menger's paper was criticized for its polemicism by Samuelson: "I myself would be a bit more sparing in use of such phrases as '...he will not, if he is sane...' [Menger(1934)], p. 212; '...behavior of normal individuals...' [Menger(1934)], p. 213; '...no normal person would risk his total fortune...' [Menger(1934)], p. 217; '...behavior of normal people...' [Menger(1934)], p. 222; '...it is clear that a normal man will risk only a limited part of his total wealth to buy chances in games' [Menger(1934)], p. 224" [Samuelson(1977)], p. 48–49. This writing style makes it difficult to detect the flaws in Menger's argument. Linguistic clarity is lost where it is urgently needed – for instance, as we shall see below, the word "risk" needs clarification, as do "sane" and "normal".

Credit is due to Arrow(1951) for writing clearly what we shall assume to be what Menger had in mind (see Section 6.2). Menger's implicit ruling-out of the fundamental ergodicity solution was perceived as "a drawback to the empirical implementation of the expected-utility criterion, since many of the most convenient forms (e.g., the logarithmic, the polynomials) seem inadmissable because of their unboundedness" [Ryan(1974)], p. 133, and Arrow commented that Menger's result "clearly meets with a good deal of resistance", [Arrow(1974)], p. 136. Nonetheless, also Arrow was misled and wrote recently, comparing to Bernoulli(1738) "... a deeper understanding was achieved only with Karl Menger's paper (1934)", [Arrow(2009)], p. 93.

Despite his stylistic criticism Samuelson had enough confidence in Menger's writing to conclude that "After 1738 nothing earthshaking was added to the findings of Daniel Bernoulli and his contemporaries until the quantum jump in analysis provided by Karl Menger", [Samuelson(1977)] p. 32, and further opined that "Menger 1934 is a modern classic that stands above all criticism", p. 49. Several renowned economists explicitly accepted Menger's argument, as reflected in passages such as "we would have to assume that U was bounded to avoid paradoxes such as those of Bernoulli and Menger" [Markowitz(1976)], p. 1278, and to my knowledge no one explicitly disagreed.

6 Errors in Menger's 1934 paper

Menger's error of conceptual type is the acceptance of an ensemble average as a relevant criterion where only a single system exists. With this acceptance he follows Bernoulli and inescapably brings the discussion to a behavioral level. D. Bernoulli explicitly states that the expectation value has to be "discarded" but he does not do so. His 1738 paper does not use the expectation value of the original gains in wealth of N. Bernoulli's work, but it still uses the expectation value of the utility of payouts $\langle \ln(D_A) \rangle$. The reader is referred to [Peters(2011)] for further details of the conceptual problem. In the following we will focus on Menger's technical errors.

6.1 The worst case in Menger's game

On p. 217 Menger(1934) introduced a modified St. Petersburg game, *game B*, whose payout for waiting time n is given by the function $D_B(n) = W \exp(2^n) - W$. Menger calls this a "slightly modified game", presumably because, like the original *game A*, it offers large payouts with small probabilities, and claims: ".. it is obvious that, even in the modified Petersburg Game, no normal person would risk his total fortune or a substantial amount", p. 217–218.

The tone of this statement motivates further probing. In this section we clarify what Menger means by the word "risk" in this sentence. We will thereby show that he probably overlooked the following

Error 1:
The worst case a player can suffer is a waiting time of $n = 1$. But the payout in this worst case is $D_B(n = 1) = W(\exp(2) - 1) \approx 6.3 \times W$. In other words, in order to risk

Table 1 The three different games used here are defined by their payout functions $D(n)$, paid if in a sequence of fair coin-tosses heads first occurs on the n^{th} toss.

Introduced by	Payout $D(n)$	Label
N. Bernoulli 1713	2^{n-1}	*A*
K. Menger 1934	$W\exp(2^n) - W$	*B*
	$\exp(2^n)$	*C*

losing anything in the game, a player has to pay for a ticket more than 6.3 times his wealth, presumably by borrowing.

Did Menger mean this by "risk", *i.e.* did Menger mean that "no normal person would" borrow as much as or more than $\$W \times (\exp(2) - 2)$ to buy a ticket, whereas a normal person may well borrow enough money to pay three times his "total fortune"? It seems that he did not mean this, as we will see. It is most likely that Menger overlooked the fact that the worst-case outcome in his game still poses no risk to the person who pays $\$W$ for a ticket.

In the original *game A* there is little difference between "risking" and "paying as a ticket price" because the difference between the worst-case net loss of $\$P - \$1$ (the amount at risk), and the ticket price, $\$P$, is only \$1. In Menger's *game B*, on the other hand, the difference between the worst-case net loss, $\$P - \$W(\exp(2) - 1)$, and the ticket price, $\$P$, is $\$W(\exp(2) - 1)$, *i.e.* many times the individual's wealth.

In his Section 2, on p. 212, Menger wrote, referring to the original *game A*: "If B, who has an infinitely large expectation, would be willing to pay any amount of money for the privilege of playing the game, then this behavior would conform to his infinitely large mathematical expectation [...] Not only will B not pay an infinitely high price to play the game, since this is impossible, buy he will not pay a very high price that he could afford. In any case, he will not, if he is sane, risk all or even a considerable portion of his wealth in a Petersburg game."

This seems to indicate that if individuals were willing to pay the highest price they can afford, Menger would have considered his paradox resolved. Assuming that the highest price one can afford is one's entire wealth (*i.e.* that borrowing is impossible), Menger's statement that "no normal person would risk his total fortune" would mean that by "risk" he meant "pay as a ticket price" and that he thought no one would pay his total fortune for a ticket in his *game B*. But since the worst-case outcome is a gain greater than the total initial fortune, there is no reason not to pay one's total fortune. It seems that Menger did not notice by how much he changed the original game.

This particular problem can be fixed by considering a different *game C*, that Menger did not propose, where the payout function is $D_C(n) = \exp(2^n)$. But the more significant error, discussed in Section 6.2, remains. The three different games involved in the discussion are summarized in Table 1.

6.2 Menger overlooked a second divergence

This section establishes the main finding, Menger's undiscovered mathematical error mentioned in the introduction. In his attempt to show that logarithmic utility fails to resolve the modified St. Petersburg *game B*, Menger showed on p. 217 that the expected payout in logarithmic utility is infinite,

$$\left\langle \Delta U_B^+ \right\rangle = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \ln\left(\frac{W + W\exp(2^n) - W}{W}\right) = \sum_{n=1}^{\infty} 1. \tag{5}$$

Menger continues with the statement: "it is obvious that, even in the modified Petersburg Game, no normal person would risk his total fortune or a substantial amount" [Menger(1934)], p. 217–218.

It was pointed out in Section 6.1 that depending on the meaning of the word "risk" this may be incorrect. The problem to be discussed now is that Menger does not actually use Bernoulli's *criterion iii*, (Eq. 3). Nor does he compute the expected net gain in logarithmic utility, which equals the time average growth rate under the dynamics discussed above, *criterion ii* (Eq. 2). He only considers the first term of *criterion iii*, (Eq. 3).

It seems that Menger's reasoning went as follows: The expected gain in utility at zero ticket price is infinite, wherefore any loss in utility resulting from a finite ticket price must be negligible. This reveals a linear way of thinking – it would be correct if utility were linear (in which case the problems of *criterion i* would resurface). Menger's analysis amounts to the introduction of

Criterion iv:
a gamble is worth taking, irrespective of the ticket price, if the expected utility-payout, (Eq. 5), is positive.

For instance, any real-life lottery, where a ticket is purchased and with a small probability something is won, is recommended by this criterion. See Table 2 for all criteria.

For clarity we turn to [Arrow(1951)], who summarized Menger as follows: "Let $U(x)$ be the utility derived from a given amount of money x, and suppose that $U(x)$ increases indefinitely as x increases [e.g. if $U(x) = \log(x)$, as in Bernoulli's own theory]. Then, for every integer n, there is an amount of money x_n, such that $U(x_n) = 2^n$. Consider the modified St. Petersburg game in which the payment when a head occurs for the first time on the nth toss is x_n. Then, clearly, the expected utility will be the same as the expected money payment in the original formulation and is therefore infinite. Hence, the player would be willing to pay any finite amount of money for the privilege of playing the game. This is rejected as contrary to intuition."

Samuelson agrees: "...Menger shows one can always fabricate a Super-Petersburg game [...] to make the compensation Paul needs again be infinite." [Samuelson(1977)], p. 32.

Both Arrow's and Samuelson's phrasings of Menger's central result agree with my reading of Menger's statements. These statements are incorrect because of

Table 2 The straw-man *criterion i* considers the expected net change in wealth. *criterion ii* considers time-average growth rates which are mathematically identical to expected net changes in utility. It resolves games *A*, *B* and *C*: there is a finite ticket price at which it discourages participation. While the utility interpretation of *criterion ii* is purely behavioral and can only be judged by experiment, the time interpretation allows predictions and judgments of when this resolution does not apply in reality. The assumption of games played at an arbitrarily high frequency becomes crucial here and is not satisfied in real life. *Criterion iii* is precisely D. Bernoulli's resolution and implies the recommendation never to pay one's entire wealth for a ticket. Menger's *criterion iv* shows that he only computed the first term of Bernoulli's *criterion iii*. Since *criterion iv* does not include the ticket price, it fails for all three games.

Introduced by	Criterion	Games resolved	Label
Huygens 1657	$\underbrace{\langle D(n)\rangle - P}_{\text{expected net wealth change}}$	none	*i*
Laplace 1814	$\underbrace{\langle \ln(W + D(n) - P)\rangle - \ln(W)}_{\text{expected net utility change = time ave. growth rate}}$	*A, B, C*	*ii*
D. Bernoulli 1738	$\underbrace{\langle \ln(W + D(n))\rangle - \ln(W)}_{\text{expected utility gain}} - \underbrace{[\ln(W) - \ln(W - P)]}_{\text{utility loss at purchase}}$	*A, B, C*	*iii*
Menger 1934	$\underbrace{\langle \ln(W + D(n))\rangle - \ln(W)}_{\text{expected utility gain}}$	none	*iv*

Error 2:

Menger overlooked that the logarithm $\ln(W-P)$ *diverges negatively at a finite value of P.*

If he did not overlook this divergence, then he overlooked Bernoulli's second term altogether. While it is correct that the first term of (Eq. 3) diverges positively, making the game appear attractive, the second term diverges negatively as $P \to W$. Samuelson's statement, paraphrasing Menger, is therefore incorrect, and Arrow's is correct only up to "Hence, the player would be willing..."

Menger's game produces a case of competing infinities. For values $P \geq W$, Bernoulli's *criterion iii* is not defined. We have to compare the divergences that lead to the infinities to understand what this undefined region signifies. To this end we consider a finite lottery, identical to *game B* except that if heads does not occur in $n_{\max}$ coin tosses, the game ends, and the lottery returns the ticket price[1]. For any finite value $n_{\max}$, the corresponding partial sum can be compensated for by a ticket price close enough to the wealth of a person, so that *criterion iii* does not favor participation in the lottery

$$\forall n_{\max} < \infty \;\; \exists \;\; P_0 < W : \sum_{n=1}^{n_{\max}} \left(\frac{1}{2}\right)^n 2^n - \ln\left(\frac{W}{W-P_0}\right) = 0. \tag{6}$$

To ensure that the criterion remains non-negative (recommending participation) up to exactly $P_0 = W$, events of zero probability must be taken into account. It is in this

[1] Laplace introduced an $n_{\max}$ in *game A*, but chose to end the lottery and return nothing after $n_{\max}$ successive tails events[Laplace(1814)], p. 440.

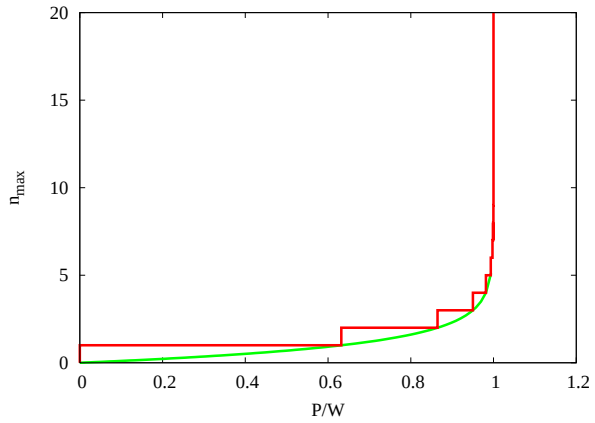

Fig. 1 Mininum (integer) number of coin tosses, $n_{\max}$, that need to be allowed in *game B* to render D. Bernoulli's *criterion iii* positive for a given price (red line), and the real number (green line) that renders the criterion exactly zero. The limiting value is $P_0 = W$, where an infinity of tosses is needed. Menger's *game B* is very profitable – even if $P_0 = W/2$ and the number of coin tosses is limited to one, the cautious *criterion iii* recommends playing.

sense that the region where (Eq. 3) is undefined corresponds to the recommendation not to buy a ticket, and the diverging expectation value of the utility change resulting from the payout is dominated by the negatively diverging utility change from the purchase of the ticket.

The dominance of the divergence of the logarithm over the diverging sum can be expressed mathematically in many different ways. For instance, (Eq. 6) can be re-written as

$$n_{\max}(P_0) = \ln\left(\frac{W}{W-P_0}\right), \tag{7}$$

implying that the maximum allowed number of coin tosses, $n_{\max}$, has to approach infinity in order to counter-balance a price close to the finite value W, see Fig. 1. Inverting (Eq. 7) yields $P_0(n_{\max}) = W\,(1-\exp(-n_{\max}))$: no matter how many terms, or coin-tosses, are taken into account, the ticket price never has to be greater than W to compensate for the diverging sum.

Far from suggesting recklessly risky behavior as *criterion i* does, Bernoulli's *criterion iii* can be criticized for being too cautious. As we have seen, the worst-case outcome of Menger's game (with $n_{\max} = \infty$) is a payout of some 630% of the player's initial wealth. At a ticket price of twice the player's wealth, the worst-case net result is still an increase in wealth of 430%, but Bernoulli's *criterion iii* discourages participating.

7 Menger's game and ergodicity

In *game B* the difference between *criterion ii* and *criterion iii* becomes visible. While Bernoulli's *criterion iii* recommends buying a ticket as long as it is less expensive than one's total wealth ($P < W$), *criterion ii* suggests to buy a ticket as long as it is not possible to lose one's entire wealth ($P < W\exp(2)$).

Criterion ii is in this sense much riskier than Bernoulli's, which raises the issue of the relationship between mathematics and reality. Is it really advisable to pay for a ticket $\$W\exp(2) - \$\varepsilon$, where ε is an arbitrarily small number? With 50% probability heads occurs on the first toss, and the player ends up penniless (except for $\$\varepsilon$). Mathematical statements are correct in a sense that statements of other type are not. The price we pay for the certainty inherent in mathematical statements is that they refer to objects that don't exist in an everyday sense of the word "exist". It would be naïve to use any mathematical criterion as a guide for decisions without questioning the criterion's conceptual meaning and the axioms upon which its derivation rests. Applicability, or relevance to the real world, pertains only insofar as the axioms and their consequences are a reflection of the real world. In the case of the expectation-value *criterion i*, this ceases to be the case when fluctuations become significant and effects of time, of events occuring in sequence, can no longer be ignored. The ergodicity resolution, *criterion ii*, assumes that equivalent lotteries can be played arbitrarily often, and this breaks down in Menger's game – games of his type are atypical in the real world. If we come across such a game, it is a once-in-a-lifetime opportunity. Mathematics cannot offer much help in this situation. It makes sense of random events only by embedding them in an ensemble of similar events, which may reside in parallel systems or within time. In the single-event setup, mathematics must shrug its shoulders and admit defeat in the face of a moral decision.

There is nothing wrong with the mathematics in the ergodicity resolution [Peters(2011)]. If we are really offered Menger-type games to be played at an arbitrarily high frequency, so that we can play arbitrarily often in our finite lives, then *criterion ii* will be practically meaningful. It is important to note that the time average growth rate (or expected net change in logarithmic utility) in *game B* does not fail to resolve Menger's paradox mathematically (unlike the expectation value in the original *game A*). There is a finite price, $P = W\exp(2)$, from which *criterion ii* discourages playing *game B*.

With payouts $D_C(n) = \exp(2^n)$ (*game C*) Bernoulli's *criterion iii* recommends to abstain from the gamble for ticket prices $P \geq W$. *criterion ii* recommends to abstain if paying the ticket price poses the risk of bankruptcy, *i.e.* if $P \geq W + \exp(2)$. *Criteria i* and *iv* fail for *game C* in the sense that *criterion i* recommends paying any finite price, though not necessarily an infinite price, and *criterion iv* recommends paying even an infinite price.

8 Summary

In 1738 D. Bernoulli made an error, or approximation – insignificant in his context – when he computed the expected net change in logarithmic utility. Perceiving this as an

error, Laplace in 1814 and later researchers corrected it implicitly without mention. In 1934 Menger unwittingly re-introduced Bernoulli's error and introduced a new error by neglecting a diverging term. Throughout the twentieth century, Menger's incorrect conclusions were accepted by prominent economists such as Arrow, Samuelson and Markowitz, although at least Arrow and Ryan noticed and struggled with detrimental consequences of the (undetected) error for the developing foramlism. The equivalence of logarithmic utility and time-averaging in the non-ergodic system allows a physical interpretation. This provides a basis for the intuition that led to the discovery of Menger's error, as discussed above.

The discussion surrounding Menger's 1934 paper deals with 3 different games, defined by their payouts $D(n)$, Table 1; and four different criteria to evaluate them, Table 2.

Criterion i

As noted by N. Bernoulli, use of the expectation-value *criterion i* produces nonsensical recommendations in *game A*, which is true also for *games B* and *C*. It is important to remember that there is no *a priori* reason for the expectation value to be a meaningful quantity because it is conceptually based on an ensemble of systems, whereas the paradox deals with only one system. The failure of *criterion i* is both mathematical (there is no finite price that should not be paid for a ticket) and conceptual (real people do not behave this way).

Criterion ii

Criterion ii expresses the irreversibility of time and the non-ergodicity of the system, and it can be phrased mathematically identically to the requirement of an expected net increase in logarithmic utility. It resolves *game A* similarly to *criterion iii*. *Games B* and *C* are mathematically resolved: no price $P \geq D(n=1)+W$ should be paid for a ticket. Behaviorally, the situation is more complicated. The assumption of infinite repeatability becomes important, and this is not usually realistic. Real people would thus be ill-advised to take this criterion literally in *games B* and *C*. Since the criterion is rooted in ergodic theory, or in the physical argument of time-irreversibility, it does not suffer from the arbitrariness of *criterion iii*.

Criterion iii

Menger claimed that Bernoulli's original *criterion iii* fails in the same sense in the modified *game B*. But that is incorrect. *Criterion iii* fails neither mathematically (no $P \geq W$ should be paid for a ticket) nor behaviorally (real people may well be limited by a no-borrowing constraint, or unwilling to pay more than their wealths). The paradox with *game C* is analogously resolved. *Criterion iii* may be criticized for being too strict – there is no fundamental reason not to pay more than $\$W$ if the minimum payout $\$D(n=1) > \$W$. It may also be criticized for being somewhat arbitrary – why is a guaranteed return of 540% acceptable but not one of 520%?

Criterion iv

Menger's *criterion iv* fails to resolve all games, even the original *game A*, because it does not include the ticket price, which appears in the second term of *criterion iii*, (Eq. 3). Irrespective of the ticket price, any gamble with a positive expected payout (not net-payout) should be taken, and no gamble with a negative expected payout should be taken (even if one were paid for participating). Menger was certainly under the impression that he was using Bernoulli's *criterion iii*, although it is unclear

whether he was aware that this is different from *criterion ii*. By using *criterion iv* Menger misrepresents Bernoulli, and his conclusions regarding both *criteria ii* and *iii* are mathematically incorrect. His rejection of unbounded utility functions is mathematically unfounded, based on errors. This is important because different utility functions can be shown to correspond to different types of dynamics. The logarithm (an unbounded function) corresponds to exponential growth, one of the simplest dynamics and a good model for many natural processes.

Acknowledgements I thank M. Gell-Mann for helpful comments and discussions. ZONlab Ltd. is acknowledged for support.

References

[Arrow(1951)] Arrow K (1951) Alternative approaches to the theory of choice in risk-taking situations. Econometrica 19(4):404–437

[Arrow(1974)] Arrow K (1974) The use of unbounded utility functions in expected-utility maximization: Response. Q J Econ 88(1):136–138

[Arrow(2009)] Arrow K (2009) A note on uncertainty and discounting in models of economic growth. J Risk Uncertainty 38:87–94, DOI 10.1007/s11166-009-9065-1

[Bernoulli(1738)] Bernoulli D (1738) Specimen Theoriae Novae de Mensura Sortis. Translation "Exposition of a new theory on the measurement of risk" by L. Sommer (1954). Econometrica 22(1):23–36, URL http://www.jstor.org/stable/1909829

[Boltzmann(1871)] Boltzmann L (1871) Einige allgemeine Sätze über Wärmegleichgewicht. Wiener Berichte 63:679–711

[Breiman(1961)] Breiman L (1961) Optimal gambling systems for favorable games. In: Fourth Berkeley Symposium, pp 65–78

[Chernoff and Moses(1959)] Chernoff H, Moses LE (1959) Elementary Decision Theory. John Wiley & Sons

[Cohen(1997)] Cohen EGD (1997) Boltzmann and statistical mechanics. In: Boltzmann's Legacy 150 Years After His Birth, Atti dei Convegni Lincei, vol 131, Accademia Nazionale dei Lincei, Rome, pp 9–23, URL http://arXiv.org/abs/cond-mat/9608054v2

[Gell-Mann and Tsallis(2004)] Gell-Mann M, Tsallis C (eds) (2004) Nonextensive Entropy: Interdisciplinary Applications. Oxford University Press

[Hanel et al(2011)Hanel, Thurner, and Gell-Mann] Hanel R, Thurner S, Gell-Mann M (2011) Generalized entropies and the transformation group of superstatistics. Proc Nat Ac Sci 108(16):6390–6394

[Huygens(1657)] Huygens C (1657) De ratiociniis in ludo aleae. (On reckoning at Games of Chance). T. Woodward, London

[Kelly Jr.(1956)] Kelly Jr JL (1956) A new interpretation of information rate. Bell Sys Tech J 35(4)

[Laplace(1814)] Laplace PS (1814) Théorie analytique des probabilités, 2nd edn. Paris, Ve. Courcier

[Markowitz(1976)] Markowitz HM (1976) Investment for the long run: New evidence for an old rule. J Fin 31(5):1273–1286

[Menger(1934)] Menger K (1934) Das Unsicherheitsmoment in der Wertlehre. J Econ 5(4):459–485, DOI 10.1007/BF01311578, URL http://dx.doi.org/10.1007/BF01311578

[Menger(1967)] Menger K (1967) The role of uncertainty in economics. English translation by W. Schoellkopf and W. G. Mellon. In: Shubik M (ed) Essays in mathematical economics in honor of Oskar Morgenstern, Princeton University Press, chap 16, pp 211–231

[Montmort(1713)] Montmort PR (1713) Essay d'analyse sur les jeux de hazard, 2nd edn. Jacque Quillau, Paris. Reprinted by the American Mathematical Society, 2006

[Peters(2010)] Peters O (2010) Optimal leverage from non-ergodicity. iFirst, Quant Fin DOI 10.1080/14697688.2010.513338

[Peters(2011)] Peters O (2011) The time resolution of the St. Petersburg paradox. In press, Phil Trans R Soc A, arXiv:10114404

[Peters and Adamou(2011)] Peters O, Adamou A (2011) Stochastic market efficiency. arXiv:11014548

[Ryan(1974)] Ryan TM (1974) The use of unbounded utility functions in expected-utility maximization: Comment. Quart J Econ 88(1):133–135

[Samuelson(1977)] Samuelson P (1977) St. Petersburg paradoxes: Defanged, dissected, and historically described. J Econ Lit 15(1):24–55

[Samuelson(1979)] Samuelson PA (1979) Why we should not make mean log of wealth big though years to act are long. J Banking Fin 3:305–307

[Samuelson(1983)] Samuelson PA (1983) Foundations of economic analysis, enlarged edn. Harvard University Press

[Todhunter(1865)] Todhunter I (1865) A history of the mathematical theory of probability. Macmillan & Co.