Title
The St. Petersburg Paradox: A Subjective Probability Solution

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Authors
Wang, Hongbin
Sun, Yanlong
Smith, Jack W.

Publication Date
2016

Peer reviewed

The St. Petersburg Paradox: A Subjective Probability Solution

Hongbin Wang (hwang@tamhsc.edu)
Yanlong Sun (ysun@tamhsc.edu)
Jack W. Smith (jwsmith@tamhsc.edu)
Texas A&M University Health Science Center
2121 West Holcombe Blvd, Suite 1109, Houston, TX 77030 USA

Abstract

The St. Petersburg Paradox (SPP), where people are willing to pay only a modest amount for a lottery with infinite expected gain, has been a famous showcase of human (ir)rationality. Since inception multiple solutions have been proposed, including the influential expected utility theory. Criticisms remain due to the lack of a priori justification for the utility function. Here we report a new solution to the long-standing paradox, which focuses on the probability weighting component (rather than the value/utility component) in calculating the expected value of the game. We show that a new Additional Transition Time (AT) based measure, motivated by both physics and psychology, can naturally lead to a converging expected value and therefore solve the paradox.

Keywords: human judgment and decision making, probability, St. Petersburg Paradox,

Fate laughs at probabilities.
-- E. G. Bulwer-Lytton

Introduction

Suppose you are offered the following gamble:

- Toss a fair coin. If you get a head, you are paid \$1 and the game is over. Otherwise, toss again.
- If you get a head in the second tossing, you are paid \$2 and the game is over. Otherwise, toss again.
- If you get a head in the third tossing, you are paid \$4 and the game is over. Otherwise, toss again.
- … Game continues until you get a head. If you get a head in the nth tossing, you will be paid $\$2^{n-1}$.

How much are you willing to pay to play this gamble?

A simple calculation shows that the gamble's expected value, S, is infinite:

$$S = \$\sum_{n=1}^{\infty} p^n 2^{n-1} = \$\sum_{n=1}^{\infty} (\frac{1}{2})^n 2^{n-1} = \$(\frac{1}{2} + \frac{1}{2} + ...) \qquad \text{Eq. 1}$$

where n is the number of tosses to get the first head (i.e., after a steak of n-1 tails, one gets a head, and the game is over).

The question is, are you willing to pay any price for a right to play this game? Probably not. More than three hundred years ago, in 1713, Nicolas Bernoulli, a young Swiss mathematician, first proposed this problem and pointed out that a sensible person would only be willing to pay very little to play the game. This constitutes a contradiction, which nowadays is called the St. Petersburg Paradox (SPP).

A Little History

The SPP was so named after the eponymous Russian city, where Daniel Bernoulli, a mathematician and Nicholas Bernoulli's cousin, published his classical solution to the problem in 1738. However, the problem was initially proposed by Nicolas Bernoulli in 1713, who was clearly troubled by it. According to him, while the expectation of game gain was infinity, the player would be guaranteed to lose since it is "morally impossible" that one not achieve a head in a finite number of tossing.

In 1728, Gabriel Cramer, another Swiss mathematician, wrote to N. Bernoulli and suggested a solution. In Cramer's solution, money's quantity was replaced by its "moral value", representing the pleasure or sorrow money (or loss of money) could produce. In doing so Cramer showed the expectation would converge to less than \$3 if "one wishes to suppose that the moral value of goods was as the square root of the mathematical quantities".

N. Bernoulli was not entirely satisfied with this solution. In his reply to Cramer, N. Bernoulli wrote that the pleasure difference "does not demonstrate the true reason" for why one should not pay infinity to play the game. Even Cramer himself thought his square-root assumption about money and pleasure was not just.

Eventually in 1738, D. Bernoulli published his solution to the problem (Bernoulli, 1738). D. Bernoulli's solution was similar to Cramer's and based on the concept of utility, which measured the usefulness of values and was taken to be a logarithmic function of values. It was shown that while the expected value diverged the expected utility converged. D. Bernoulli's solution was seminal and extremely influential, and has since shaped the whole field of economics and of the psychology of decision making.

It was interesting to note that N. Bernoulli vigorously objected his cousin's approach. A series of communication showed that the two had engaged in serious arguments. To N. Bernoulli, the concept of utility, similar to the "moral value" of Cramer, was arbitrary and, to a certain extent, irrelevant. Rather, the concern here was to find a more general way to show if a game was fair, regardless of who was playing the game. "For example a game is considered fair, when the two players bet an equal sum on a game under equal conditions, although according to your theory, and by paying attention to their riches, the pleasure or the advantage of gain in the favorable case is not equal to the sorrow or the disadvantage that one suffers in the contrary case".

N. Bernoulli had his own insights on how to solve the problem. To him, the true reason had to do with those "cases which have a very small probability, [which] must be neglected and counted for nulls, although they can give a very great expectation". Unfortunately, N. Bernoulli encountered great difficulty in deciding when a very small probability should be counted as zero. As a result, his approach was not fully developed at that time and was completely overshadowed by the utility-based solutions.

In this paper we argue that N. Bernoulli might be correct and we provide a new and complete mathematical treatment that is consistent with his insights. Before we dive in, however, we would like to briefly review existing solutions to the St. Petersburg Paradox.

The St. Petersburg Paradox

The key puzzle behind the SPP is how a gamble with an infinite expected return could be valued so little to a human player. Over 300 years many solutions have been proposed, which can roughly be divided to three categories.

The first line of solution attacks the realism of the gamble. Given that the expected gain for one player is infinite, the potential loss for the other player would be infinite as well. Since nobody has infinite payout, the game in its classical form is not realistic and will in no way be offered in the real world. Therefore, if we cap the potential payout in a revised game, the expected value would then converge.

Even D. Bernoulli had thought this issue was important. In 1731 he wrote to his cousin, "I have no more to say to you, if you do not believe that it is necessary to know the sum that the other is in position to pay". The wiki page on SPP (http://en.wikipedia.org/wiki/St._Petersburg_paradox) has a table listing the expected values with respect to a few interesting caps. For example, if the payout is capped with the total US GDP in 2007 (~$13.8 trillion), the expectation is merely $44.57.

The second category of solution focuses on the value component of Eq. 1. The argument is that people are not interested in the monetary value per se, but more in the utility, goodness, or pleasure the money brings forth, which can be represented by a utility function. Therefore, if the value is replaced by a utility function, which is typically assumed to be concave, it can be shown that the expectation would converge. Cramer's and D. Bernoulli's solutions belong to this category.

Utility is capable of capturing the time discounted value of wealth, and can simultaneously incorporate different risk preferences. The utility theoretical approach has since become a dominant solution to the game and has enjoyed a profound influence in the broad field of economics (von Neumann & Morgenstern, 1944). However, criticisms regarding the correct form of utility function remain. Note that both the square-root function and the logarithmic function, as well as many other concave functions, would work (in that the total sum converges). Should one solution be preferred to another? D. Bernoulli tended to believe the distinction was not important. He praised Cramer's solution in his paper, "Indeed I have found his theory so similar to mine that it seems miraculous that we independently reached such close agreement on this sort of subject" (Bernoulli, 1738).

Whereas it appears that it is an empirical issue and can only be answered by psychological investigations, Ole Peters, a physicist, recently proved, via a solid mathematical treatment, that the logarithmic function could be naturally derived based on the mathematics of time average of rate of return and was therefore necessary in the situation (Peters, 2011). Peters argued that D. Bernoulli accidentally chose the correct function form for utility even though he was not aware of its underlying physics. However, it is important to note that while Peter's treatment removes the arbitrariness in the utility functional form, it involves additional assumptions not in the original SPP such as repeated games (Buchanan, 2013).

The third line of solution focuses on the probability weighting component of Eq. 1 and argues that people simply regard those extremely large payout cases having so small a probability that render them impossible to occur. As we mentioned above, N. Bernoulli was first advocate of this idea. In a 1728 letter to Cramer, he wrote,

"with him a very small probability to win a great sum does not counterbalance a very great probability to lose a small sum, he regards the event of the first case as impossible, and the event of the second as certain. It is necessary therefore, in order to settle the equivalent justly, to determine as far as where the quality of a probability must diminish, so that it be able to counted null".

With this reasoning, he demonstrated, for example, that if one regarded probabilities less than 1/32 as null, then the expectation became merely $2. However, he recognized his difficulty and got stuck. He continued to write,

"but here is what which is impossible to determine, any assumption that one makes, one encounters always difficulties; the limits of these small probabilities are not precise".

On Subjective Probability of Rare Events

N. Bernoulli's difficulty has to do with the lack of a theory on how the human mind represents and processes rare events with very small probabilities. While it is certainly unsatisfying to set an arbitrary limit so that any smaller probabilities are treated as zero, it is important to recognize that the way in which the human mind represents probabilities may not be same as what the standard probability theory prescribes.

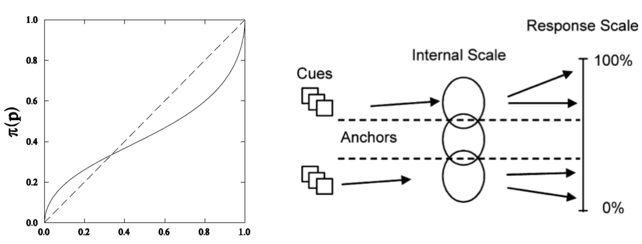
At human scales, very small probabilities are often linked to impossibility. Emile Borel, an eminent French mathematician, introduced what he called the single law of chance in his 1943 book. The law, which nowadays is simply called Borel's Law, says, "Events with a sufficiently small probability never occur". Borel goes on to clarify what he meant by "sufficiently small" probabilities, by distinguishing different scales, as shown in Table 1. According to Borel, therefore, probabilities smaller than 1 in a million should be regarded as small enough that for practical purposes a human can treat them as impossible. In the context of SPP, one in a million roughly corresponds to getting 20 tails in a row. By setting smaller probabilities to zero, the expectation for the gamble reduces to about \$10.

Table 1: Negligible probabilities according to Borel's law

Scale	Limit	Example
Human	1 in 10^6	Any two chosen seconds in a year are same
Terrestrial	1 in 10^{15}	Any two chosen square foot on Earth are same
Cosmic	1 in 10^{50}	Any two chose atoms on Earth are same
supercosmic	1 in $10^{1,000,000,000}$	Any two particles in the universe are same

That people distort probabilities from their standard values in a systematic way is also a critical claim in prospect theory, a descriptive theory of human decision making (Kahneman & Tversky, 1979; Tversky & Kahneman, 1992). According to this theory, utilities should be weighted by the psychological correspondence of probabilities rather than the standard probabilities themselves, and the relationship between the two can be represented by a π function, as shown in Figure 1. Therefore, it seems that people are systematically underestimating large probabilities and overestimating small probabilities.

The π function has been used to explain many interesting phenomena such as the certainty effect. However, when applied in the SPP situation it works in the opposite direction - here small probabilities need to be further reduced rather than enlarged. Nevertheless, the π function is consistent with the general idea that subjective probabilities are not necessarily equal to objective probabilities.



A B

Figure 1: (A) The π function as described in prospect theory (based on Tversky and Kahneman, 1972); (B) Internal representations on a coarser scale have to be transformed to external probabilities on a finer scale, leading to distortion.

Such distortion, despite the difference in specific forms, may result from the limited resolution of the mind in representing uncertainty. We (Sun, Wang, Zhang & Smith, 2008) provided evidence supporting that the mind may adopt a probability scale much coarser than the one prescribed in probability theory. For example, the difference between p=0.0125 and p=0.00625 may be significant in probability theory (and in the standard SPP context), they may all be represented as "quite small", "unlikely", or "impossible" in the mind. This lack of resolution, in addition to other constraints such as anchor and adjustment, contributes to a distorted mental representational scheme of probabilities (see Figure 1B).

Probability and Time

It is desirable to develop a normative mathematical treatment that at the same time describes how the human mind perceives and represents probabilities. Hopefully, such a treatment can naturally lead to a solution to the SPP. Among other advantages, this treatment is consistent with N. Bernoulli's original insights to the problem when he first proposed it, and avoids the inherent arbitrariness of the utility-based approach.

In recent years we have advocated an approach that is based on the inherent connection between probability and time (Sun & Wang, 2010a; 2010b; 2013; Sun et al, 2015). A comprehensive treatment of the connection is still under development. In this paper we provide a concise and relevant narrative with a goal to demonstrate how it can be applied to solve the SPP.

Although the concept of probability is often formally defined using a set theoretical axiom system (e.g. Kolmogorov, 1965), there exists an intriguing relationship between probability and time. In general, for a given event, its probability describes the relative frequency of its occurrence in the long run. In the context of time, the probability of an event corresponds to the mean inter-arrival time (MT) of the event, describing how long it takes for the same event to occur again.

To facilitate, consider the situation of fair coin tosses, where Head (H) or Tail (T) can occur in each toss. We know,

P(T)=1/2, P(TT)=1/4, P(TTT)=1/8, P(TTTT)=1/16.

This is equivalent to say, in terms of the MT:

MT(T)=2, MT(TT)=4, MT(TTT)=8, MT(TTTT)=16.

That is, on average it takes 2 tosses for T to re-appear, 4 tosses for TT (2 Tails in a row) to re-appear, 8 tosses for TTT to re-appear, 16 tosses for TTTT to re-appear. In fact, it can be proved that the standard probability is simply the reciprocal of mean time:

$$p = 1 / MT \qquad \text{Eq. 2}$$

It turns out that there is another time statistic that is also associated with uncertain events, waiting time (WT). WT describes how long one has to wait for an event to occur for the first time (rather than to re-occur). Intuitively, one would think a more frequent event (i.e., an event with a short mean time) would occur soon (i.e., with a short waiting time), however, it is not generally so. The MT and WT can be dissociated, especially for random sequences with length longer than 1. In particular, it can be shown that streak patterns have the longest WT among sequences with equal lengths:

WT(T)=2, WT(TT)=6, WT(TTT)=14, WT(TTTT)=30.

That is, if one starts to toss a fair coin, on average it takes 2, 6, 14, and 30 tosses for T, TT, TTT, and TTTT to occur for the first time, respectively. This is different from these patterns' MTs (2, 4, 8, and 16, respectively). More complete treatments of WT can be found here (Sun & Wang, 2010a; 2010b; 2013).

An intuitive explanation for why the WT for streak patterns is the longest is that when a streak is interrupted, it takes longer to get back to that streak. Therefore, given a random sequence, streaks have larger variances than non-streaks. This statistical fact leads to another time-based statistic, which we call Additional Transition Time (AT).

Formally, a flipped coin can be viewed as a Bernoulli process with the probability of heads p_H (or probability of tails p_T). We can then associate it with a Markov chain, whose states are patterns consisting of outcomes generated by consecutive tosses (Figure 2). We consider the AT, $A_{i,j}$, as the transition time (i.e., the number of tosses it takes in this case) for the Markov chain to first reach pattern j, given the current pattern i.

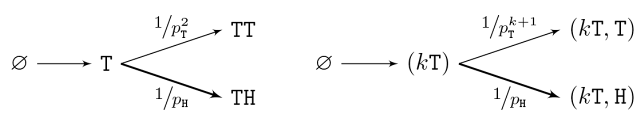


Figure 2: A binomial tree representing the expected AT when an existing streak is either continued or discontinued by a single additional trial. ϕ represents the very beginning of the process (i.e., start anew with a time stamp of zero). The left figure is a special case of the right figure with $k = 1$. For a fair coin $p_H = p_T = 1/2$, given the current state T, the process has the same probability to branch into either TT (continuation) or TH (discontinuation). However, from time's perspective, TH is more immediate (2 tosses away) than TT (4 tosses away). As the length of the initial streak k increases, the AT remains the same for a discontinuation of the streak ($1/p_H$), but grows exponentially for a continuation of the streak ($1/p_T^{k+1}$).

Figure 2 illustrates a quite striking result. Given that we have observed TTT, for example, how much additional transition time it takes to reach TTTT? Standard probability theory tells us there is an equal probability (1/2) of H or T in the fourth toss. However, on average, it would take $1/(1/2)=2$ additional tosses to get TTTH, and $1/(1/2^4)=16$ additional tosses to get TTTT – *TTTT has a much longer AT*. Again, a simple explanation is that in waiting for the streak, if it goes awry, the wait would have to re-start from scratch. This is in contrast to TTTH, where a wrong outcome (i.e., expecting a final H but getting T) will not hurt much as the waiting (of a final H) can continue. Taken together, this is just another mathematical fact that justifies why streaks are rare and remarkable and why (kT,H) is more imminent than (kT,T).

It can be shown that the AT is just another manifestation of the waiting time. Both statistics are affected by the same start-anew effect due to the self-overlapping property of the streak pattern. Instead of treating each pattern as a whole (as in calculating the waiting time), the AT allows temporal prediction by breaking the waiting time statistics into two parts, as follows,

$$\begin{aligned} WT(kT,H) &= WT(kT) + MT(H) = WT(kT) + 1/p_H \\ WT(kT,T) &= WT(kT) + MT(kT,T) = WT(kT) + 1/p_T^{k+1} \end{aligned} \qquad \text{Eq. 3}$$

Thus, given a streak of k tails, the expected AT for the streak to be extended by one more tail is not the mean time of a single tail (MT(T)), but the mean time of k+1 tails (MT(kT,T)).

While it appears counter-intuitive, WT and AT capture an essential environmental statistic describing when an event is to occur. They are certainly relevant to human cognition. In many everyday situations, it is likely that the question of when an event is to occur is more important than the question of how often an event is to occur. Therefore, it is plausible that the brain and the mind have developed mechanisms to be sensitive to WT and AT statistics. We have previously argued that human perception of randomness in general and the gambler's fallacy in particular might be linked to the longer WT of streak patterns (Sun & Wang, 2010a; 2010b; 2013). More recently, we have shown how the brain could learn to capture the WT statistic through predictive neural learning (Sun et al., 2015).

Toward an AT Based Solution to SPP

It is therefore quite plausible that the AT captures, both normatively and descriptively, people's sensitivity to the rarity of streak patterns. In this account, TTTT is different from patterns such as TTTH not only in terms of the additional transition time it requires to complete the pattern, but also the fact that the difference increases quickly as the pattern length increases.

The SPP inherently involves streak patterns. For the player to win big, he/she would wish to delay the first occurrence of H as much as possible, that is, to get the streak of Ts as long as possible. According to N. Bernoulli, it is these long streaks that should be rendered as impossible to ever occur due to their very small probabilities. Unfortunately, the standard mean time based probability theory does not distinguish between streak and non-streak patterns and treats them equally likely – both as a function of pattern lengths. We argue that that the concept of AT offers a new perspective to resolve N. Bernoulli's difficulty and can lead to a more justified solution to the SPP.

More specifically, we have shown that having already observed TTT, while the probability of getting the fourth T is ½, the AT for getting TTTT is 16. Given the genuine relationship between probability and time, it is possible to derive another probability measure based on the AT as follows:

$$p' = 1/AT \qquad \text{Eq. 4}$$

p', therefore, measures a type of uncertainty associated with obtaining the final outcome required to complete the entire sequence. Different from *p* (Eq. 2), *p'* is pattern structure dependent – different prefixes result in different *p'*. In essence, it is the local structures of patterns that differentiate patterns and make streaks special. Mean-time based *p* is blind to the structures and AT-based *p'* highlights the difference.

We have shown that the AT it takes for a streak of length (k-1) to be extended to length k is $1/p^k$. Thus, we can derive the following relationship:

$$p' = p^k \qquad \text{Eq. 5}$$

In the example above, to get the fourth T given TTT, we have p=1/2, p'=1/16.

We can then replace the context-free *p* in Eq. 1 with the context-sensitive *p'* to acquire another expectation measure as follows:

$$S' = \$\sum_{n=1}^{\infty}\left(\prod_{k=1}^{n} p'_k\right)2^{n-1} = \$\sum_{n=1}^{\infty}\left(\prod_{k=1}^{n} (1/2)^k\right)2^{n-1} \qquad \text{Eq. 6}$$

Different from S, which diverges, it can be proved that S' converges.

Figure 3 depicts S' converging behavior as the number of trials increases. For comparison, we also plot the logarithmic utility solution by D. Bernoulli. It is clear that S' converges very fast, to the asymptotic value as early as toss 5. In a sense this solution fits N. Bernoulli's original insights almost perfectly. As mentioned above, he suggested that a sensible man should treat those cases with probability less than 1/32 as impossible, which corresponds to streaks of TTTTT and longer. However, our solution avoids N. Bernoulli's difficulty and makes an otherwise arbitrary choice justifiable. According to this solution, long streak patterns occur with probabilities that decrease so fast that they over-compensate the rather large payouts in those situations. The net expectation converges, supporting the perceived limited value of the gamble.

Another feature of the new AT-based solution is that it does not need a concave utility function to resolve the puzzle. N. Bernoulli criticized that focusing on utility rather than value was not the true reason for why people valued the gamble less. Although the logarithmic utility function has received much support in the past as a reasonable way describing how people conceive goodness of value, there is a lack of a priori justification for the choice.

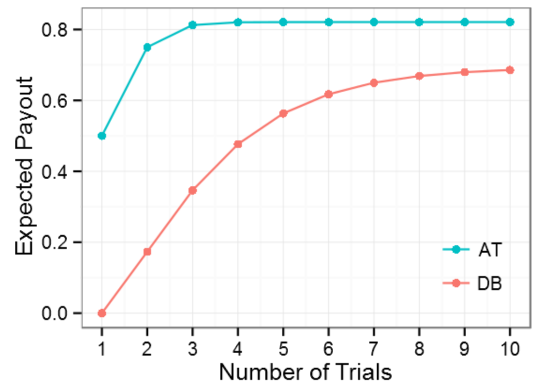


Figure 3. Converging expected payout in the SPP. **DB**: the logarithmic utility solution by Daniel Bernoulli. **AT**: new AT-based solution (Eq. 6).

Discussion

The St. Petersburg Paradox is an over 300-year old puzzle and still resides in the center of any formal understanding of human decision making. Why is a gamble with an infinite expected return valued so little? The classical solution, proposed by Daniel Bernoulli, resorts to the concept of utility. According to this account, humans do not value money at its face value, rather, its utility, measuring its usefulness or the pleasure it brings about, should be considered. Utility is apparently sensitive to, and therefore is capable of capturing the effect of, a range of factors, including individual difference, risk preferences, and time discount. The concept is so intuitively appealing and mathematically powerful that it has since become a cornerstone of modern economics. However, the lack of a priori analysis for choosing a specific form of utility function raises a problem. More recently, Ole Peters demonstrated that the logarithmic utility function is a mathematically necessary result if a probabilistic decision maker is assumed to maximize return over time.

Nicolas Bernoulli, the original proposer of the puzzle, hypothesized that the culprit was those cases that carry large payouts but with very small probabilities. He suspected that to a sensible human those cases should be rendered as impossible to occur. Equipped with standard probability theory, however, he encountered great difficulty in deciding when a probability was small enough.

In this paper we have suggested a different solution to the problem. The solution is consistent with N. Bernoulli's probability weighting idea, but avoids its difficulty. Essentially, we have derived a different probability measure based on the waiting time, an environmental statistic that describes how soon an uncertain event is to occur. It can be shown that in tossing a fair coin, while patterns of equal length all have the same mean time, that is, the same probability, they may have different waiting times. We show that via the concept of Additional Transition Time, a different probability measure can be derived for patterns. In particular, we show that streak patterns have a much longer AT than non-streak patterns, rendering them to be very rare to occur. We demonstrate that this probability weighting leads to a converged expected return and therefore solves the SPP.

One advantage of our solution is that it eliminates arbitrary ad-hoc choices, for either the utility function form or the limit for sufficiently small probabilities. It is a normative solution based on mathematics, with fewer pre-assumptions. The brain's sensitivity to the waiting time has recently been demonstrated, which lends further credibility to the solution.

In sum, the treatment presented in this paper is in contrast to almost all existing theories attempting to explain and rationalize human biases in judgment and decision making (Falk & Konold, 1997; Gigerenzer & Hoffrage, 1995; Gilovich, Griffin & Kahneman, 2002; Griffiths, Chater, Kemp, Perfors & Tenenbaum, 2010; Ma, Beck, Latham & Pouget, 2006). In spite of different details, these theories are all based on the assumption that human mind encodes mean-time based probabilities or likelihoods. Here we argue that an accurate encoding of more complicated temporal structures (specifically, the wait time and additional transition time statistics) is at the core of how people represent uncertainty.

Acknowledgments

This work was partially supported by the Air Force Office of Scientific Research (AFOSR) grant number FA9550-12-1-0457, Office of Naval Research (ONR) grant number N00014-16-1-2111.

References

Bernoulli, D. (1738). Specimen theoriae novae de mensura sortis [exposition of a new theory on the measurement of risk, translated by louise sommer, 1954, published on econometrica, 22, 23-36]. *Commentaries of the Imperial Academy of Science of Saint Petersburg.*

Borel, E. (1962). *Probabilities and life [translated from probabilités et la vie, 1943, by maurice baudin].* New York, Dover Publications.

Buchanan, M. (2013). Gamble with time. *Nature Physics*, *9*(1), 3-3.

Falk, R. & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*, *104*(2), 301-318.

Gigerenzer, G. & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684-704.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of inuitive judgment.* New York: Cambridge University Press.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences.*

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263-291.

Kolmogorov, A. N. (1965). Three approaches to the quantitative definition of information. *Problems of Information Transimission*, *1*, 1-7.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432-8.

de Montmort, P. R. (1713). *Essay d'analyse sur les jeux de hazard [essays on the analysis of games of chance] (reprinted in 2006) (in french) (second ed.).* Providence, Rhode Island: American Mathematical Society.

Peters, O. (2011). The time resolution of the st petersburg paradox. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *369*(1956), 4913-31.

Sun, Y. & Wang, H. (2010a). Gambler's fallacy, hot hand belief, and the time of patterns. *Judgment and Decision Making*, *5*(2), 124-132.

Sun, Y. & Wang, H. (2010b). Perception of randomness: On the time of streaks. *Cognitive Psychology*, *61*, 333-342.

Sun, Y. & Wang, H. (2013). Finding structure in space and time: Active trace of patterns. In *Proceedings of the 35th annual conference of the cognitive science society.* Cognitive Science Society.

Sun, Y., O'Reilly, R. C., Bhattacharyya, R., Smith, J. W., Liu, X., & Wang, H. (2015). Latent structure in random sequences drives neural learning toward a rational bias. *Proceedings of the National Academy of Sciences of the United States of America, 112(12),* 3788-3792.

Sun, Y., Wang, H., Zhang, J., & Smith, J. W. (2008). Probabilistic judgment on a coarser scale. *Cognitive Systems Research*, *9*(3), 161-172.

Tversky, A. & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk & Uncertainty*, *5*, 297-323.

von Neumann, J. & Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton, NJ: Princeton University Press.