The St Petersburg paradox

COMPSCI 3016: Computational Cognitive Science
Dan Navarro & Amy Perfors
University of Adelaide

Abstract

This note provides a somewhat cleaner discussion of the St Petersburg paradox than the lecture slides. If anything the notes doesn't make sense please contact me (these notes were written by Dan: daniel.navarro@adelaide.edu.au) and I'll try to fix them!

The Game

The St Petersburg paradox involves around a simple gambling game, in which the gambler pays a fixed cost \$x to enter the game, and then you receive some payoff \$y based on the results of the coin flips. The idea is that there is a "pot" that starts out at \$1, which doubles every time that a (perfectly fair) coin comes up heads. The game stops as soon as the coin comes up tails, and the gambler takes home the current value of the pot. For instance, if the coin flips are HHHT, then the gambler takes home a pot of $y = \$8$. On the other hand if the result is T (i.e., the first flip comes up tails), the gambler takes home the initial pot $y = \$1$. The question is: how much money x should we be willing to play in order to enter the game.

The Paradox

Let's consider this from the perspective of classical expected utility theory. Formally, the utility of playing a game that charges \$x to play can be written:

$$u(\text{game}) = \sum_{n=0}^{\infty} u(n, x) P(n) \tag{1}$$

and the utility of not playing the game is 0 (neither good nor bad). In this expression n denotes the number of heads that you observe before the first tail, and $P(n)$ denotes the probability that exactly n heads occur before the first tail. If the coin is fair, the probability of observing n heads and then a tail is $(1/2)^{n+1}$. In classical theory, the utility of paying $\$x$ to win $\$y$ is just the difference between the two: $y - x$. If you win more money than you paid, you have a gain, and if you win less money than you paid, you have a loss. The key thing is that the utility scales linearly. In the game, if there are n heads before the first tail, then the winnings are $y = 2^n$. So, having made these assumption we can calculate the utility of a game that charges an entry fee of $x:

$$
\begin{aligned}
u(\text{game}) &= \sum_{n=0}^{\infty} (2^n - x) \left(\frac{1}{2}\right)^{n+1} && (2)\\
&= -x + \sum_{n=0}^{\infty} 2^n \left(\frac{1}{2}\right)^{n+1} && (3)\\
&= -x + \sum_{n=0}^{\infty} \frac{1}{2} && (4)\\
&= -x + \infty && (5)
\end{aligned}
$$

As a result, the expected value of the game is infinite. No matter how much money x you charge for entry, the expected utility is positive, and so you should be willing to play. Yet in practice, most people are only willing to pay a few dollars to play. Why is this happening?

Solution 1: Logarithmic utility

The first solution to the problem is to alter the utility function to reflect what is probably a pretty sensible psychological insight: the subjective value of a dollar isn't fixed, and actually depends on how much money you already have. As Bernoulli put it

> The determination of the value of an item must not be based on the price, but rather on the utility it yields ... There is no doubt that a gain of one thousand ducats is more significant to the pauper than to a rich man though both gain the same amount.

One natural way to characterise this insight is to say that the utility $u(b)$ of a bank balance b scales logarithmically; i.e., $u(b) = \ln b$. Let's consider the value of a game that charges x dollars, to a player who starts the game with b dollars in the bank:

$$
\begin{aligned}
u(\text{game}) &= \sum_{n=0}^{\infty} u(n, x, b) P(n) && (6)\\
&= \sum_{n=0}^{\infty} \ln(2^n + b - x) \left(\frac{1}{2}\right)^{n+1} && (7)
\end{aligned}
$$

The utility of not playing the game, and retaining your bank balance of b dollars is $u(\text{no game}) = \ln b$. So, a player with b dollars in the bank should be willing to play a game that charges x if the difference between the two is positive. That is, you should play if:

$$u(\text{game}) - u(\text{no game}) > 0 \quad (8)$$

$$\left(\sum_{n=0}^{\infty} \frac{\ln(2^n + b - x)}{2^{n+1}} \right) - \ln b > 0 \quad (9)$$

$$\sum_{n=0}^{\infty} \frac{\ln(2^n + b - x) - \ln b}{2^{n+1}} > 0 \quad (10)$$

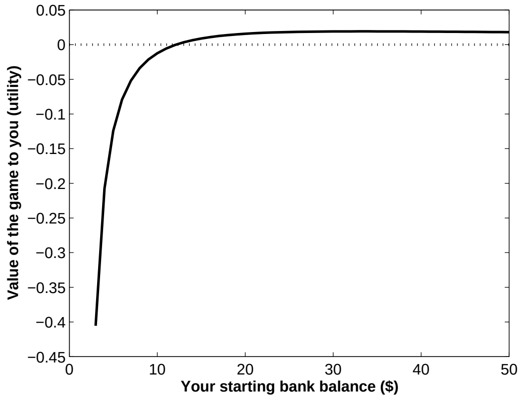
```
x = 3; % cost of the game
B = x:50; % bank balance of the player

U=zeros(size(B)); % utilities
for i=1:length(B) % over all bank balances

    b = B(i); % bank balance
    logb = log(b); % utility of not playing
    u = 0; % initialise utility

    for n = 0:1000 % over # of heads (1000 = overkill)
       u = u + (log(b - x + 2^n) - logb) / (2^(n+1));
    end
    U(i)=u;
end
```

(a)



(b)

Figure 1. MATLAB code calculating the expected utility of a game that costs \$3 as a function of the bank balance, (a) and a plot of the corresponding function.

This isn't a very pretty expression, but notice as n gets large, the numerator $\ln(2^n + b - x) - \ln b$ is approximately $\ln(2^n)$. So for large n, the summands converge to $\frac{\ln(2^n)}{2^{n+1}}$, which in turn approaches 0 for large n. The value of this function is calculated by the code in Figure 1a and plotted in Figure 1b.

Solution 2: Finite wealth of the casino

A clever alternative solution is to point out that not only do you need to consider the bank balance of the gambler, but also the bank balance of whoever is offering the bet. Let's call this other party "the casino", and for simplicity, we'll use the original linear utility function, rather than Bernoulli's logarithmic one. We let w denote the bank balance of the casino, and we note that (in the simplest version of this "finite wealth" solution) the casino can't possibly pay the gambler more money than w. If we let $n^* = 1 + \lfloor \log_2(w) \rfloor$, then notice that if the gambler obtains $n \geq n^*$ heads before the first tail appears, he or she will only get a payout of w, even though he or she is entitled to get 2^n. This is quite plausible: in real life you can't sue someone for money they don't have. In other words, the actual payout given by the game is, in practice, $\min(2^n, w)$. What this means is that the expected value of the game is now

$$
\begin{aligned}
u(\text{game}) &= \sum_{n=0}^{\infty} \frac{1}{2^{n+1}} \min(2^n, w) && (11)\\
&= \sum_{n=0}^{n^*-1} \frac{1}{2^{n+1}} 2^n + \sum_{n=n^*}^{\infty} \frac{1}{2^{n+1}} w && (12)\\
&= \sum_{n=0}^{n^*-1} \frac{1}{2} + w \sum_{n=n^*}^{\infty} \frac{1}{2^{n+1}} && (13)
\end{aligned}
$$

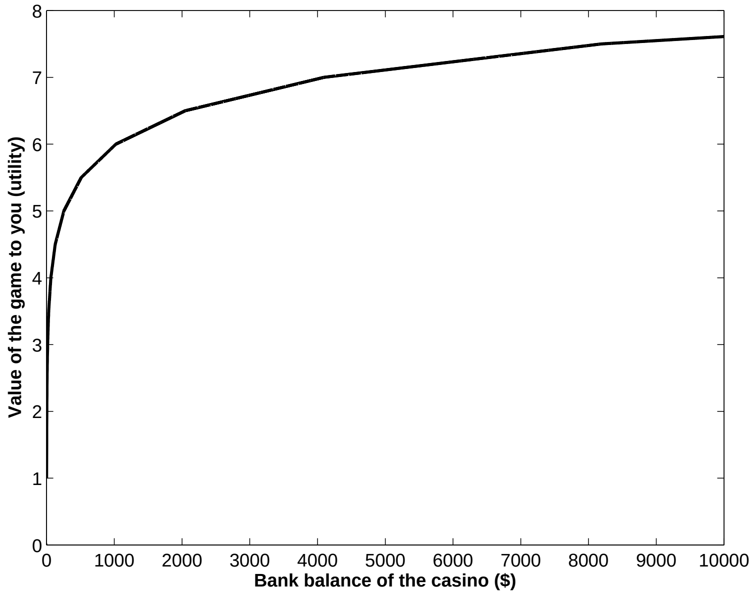


Figure 2. The expected value of the game when the wealth of the casino is finite, as per Equation 14.

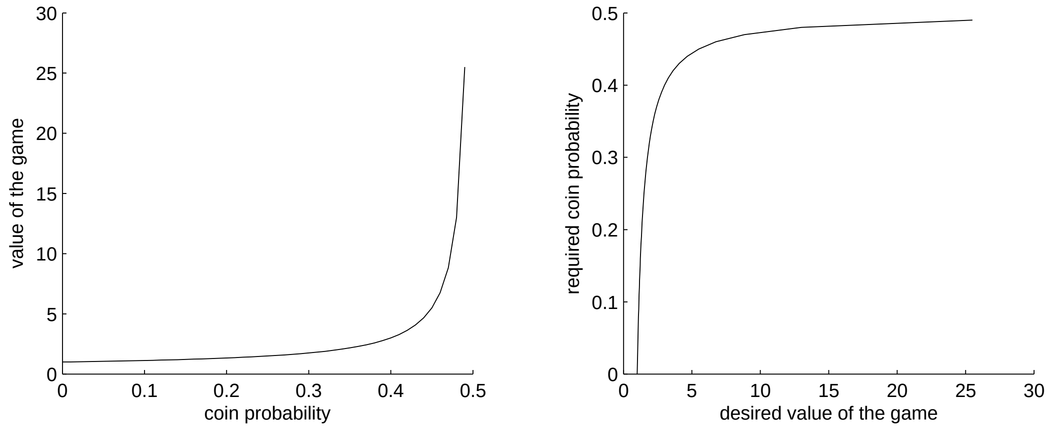
$$= \frac{n^*}{2} + \frac{w}{2^{n^*}} \quad (14)$$

Unlike Bernoulli's original formulation of the problem, this one clearly has finite expected value. The function in Equation 14 is plotted in Figure 2. In fact, even if you give the casino a bank balance of $1 billion, the expected value of the game still only rises to $15.93.

Solution 3: Slightly biased games

An interesting variation of the finite wealth argument (suggested by a student) is to argue that the casino can only legitimately offer games that have a finite expected value (otherwise there's no way they can afford to play the game, regardless of what kind of insurance they have!). As such, the question becomes: how much does the casino have to alter the game in order to be able to legitimately offer it to the gambler? The simplest way to alter the game and still keep the expected value finite is to offer a game in which the multiplier on the pot is *slightly* lower than the probability of a head. That is, if we let $\theta = P(H)$ denote the probability that the coin comes up heads in the gamble, then the casino can only offer games in which each successive head multiplies the value of the pot by some number less than θ. Let ϵ be the difference between the pot multiplier and the probability that the coin comes up heads. That is, the pot multiplies by $\theta(1 - \epsilon)$ after each heads. Then the expected value of the game is:

$$u(\text{game}) = \sum_{n=0}^{\infty} u(n)P(n) \quad (15)$$



(a) (b)

Figure 3. The relationship between the value of the game, and the underlying coin probability.

$$= \sum_{n=0}^{\infty} \left(\frac{1-\epsilon}{\theta}\right)^n \theta^n (1-\theta) \quad (16)$$

$$= (1-\theta) \sum_{n=0}^{\infty} (1-\epsilon)^n \quad (17)$$

$$= \frac{1-\theta}{\epsilon} \quad (18)$$

where the last line follows from the fact that the summation is a geometric series. If the casino intends to offer a gamble in which the pot doubles every time the coin comes up heads then we have

$$\frac{1-\epsilon}{\theta} = 2 \quad (19)$$

and thus

$$\epsilon = 1 - 2\theta \quad (20)$$

Therefore, if the casino wants the game to have an expected value of v, then they need to rig the coin to have probability θ such that

$$\frac{1-\theta}{1-2\theta} = v \quad (21)$$

This relationship between the value of the game v and the coin probability θ is plotted in Figure 3, from two different perspectives. On the left side (panel a), we can see that as the coin probability gets closer to being a fair coin (i.e, as $\theta \rightarrow 1/2$), the value of the game becomes infinite. However, from the *casino's* perspective, the plot on the right (panel b) is more interesting: if you want to keep the value of the game to a small, finite value v, you don't have to rig the coin very much. A coin set to $\theta = .49$ yields a game that has a value of \$25.50. A coin set to $\theta = .499$ yields a game worth \$250.50.