ELSEVIER

Acta Psychologica 93 (1996) 59–68

acta psychologica

Probability judgments in multi-stage problems: Experimental evidence of systematic biases

Uri Gneezy *

CentER for Economic Research, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

Abstract

We report empirical evidence that in problems of random walk with positive drift, bounded rationality leads individuals to underestimate the probability of success in the long run. In particular, individuals who were given the stage-by-stage probability distribution failed to aggregate this information in a multi-stage case. Estimations of the long-run probability distribution did not differ much from the given stage-by-stage probability distribution, and were systematically lower than the accurate one. Applications to risk perception in financial markets are considered.

PsycINFO classification: 2340

Keywords: Estimation bias; Compound probability

1. Introduction

The standard approach to sequential decision making under uncertainty (i.e. the Savage (1954) subjective expected utility theory) assumes that people are indifferent to the way problems are set, and are only interested in the probability distribution over final outcomes (see Hammond, 1988; Machina, 1989). In particular, people are assumed to follow the reduction of compound lotteries axiom, stating that a multi-stage lottery is equally attractive as the one-stage lottery that yields the same prizes with the corresponding multiplied probabilities. For example, consider the following two lotteries: in the first, a fair coin is tossed twice in a row. If it falls on its head twice or on its tail twice, the decision maker wins \$1; he loses \$1 if it falls once on each side. In the second lottery, two fair coins are tossed at the same time, and the payoffs are the same as in the

* E-mail: gneezy@kub.nl

PII S0001-6918(96)00020-0

first lottery. The reduction axiom states that the decision maker is indifferent between the first and the second lottery.

While there exists a literature following Kahneman and Tversky (1979) (see Camerer, 1995, for a short survey) that questions this axiom, a common assumption is that people are capable of accurately estimating the reduced probabilities of compound lotteries, or at least that mistakes are not systematic, and estimations are accurate on average. This is surprising because it is not difficult to construct sequential problems in which, for bounded rationality reasons, people fail to estimate reduced probabilities (see the book edited by Kahneman et al., 1982, or the works of Bar-Hillel, 1973, and Wagenaar and Sagaria, 1975, which are described below).

This paper looks at a different aspect of this problem, namely processes of random walk with positive drift, which are very important in many real-life economic decisions. We test whether probability judgment is 'good' in this kind of environment, or is it systematically biased – and if so how. Versions of the following investment game [1] are used:

> An option on the price of a stock is for sale. Today the price of the stock is $\$x$, and every day it either goes up or down by \$1, with probability p and $1-p$ respectively. The option will be realized and pay \$0 if the price of the stock will reach \$0, and $\$n$ if the price will reach $\$n$. What is the probability that the realization price will be $\$n$?

Results of three experiments with the game are reported, in which we controlled for the following three parameters:

1. we changed the starting amount to $x=\$3$, \$5, and \$7, fixing $p=0.6$ and $n=10$,
2. we changed the size of the interval to $n=4$, 6, 10, and 14, when $x=n/2$ and $p=0.6$, and
3. for $x=5$ and $n=10$, we changed the stage-by-stage probability of success to $p=0.55$, 0.58, 0.6, 0.65, and 0.7.

The results suggest that people use the stage-by-stage probability as an anchor, and adjust insufficiently. Estimations are biased toward the direction of the stage-by-stage probability, resulting in underestimation of the overall probability of success. One consequence is that while individuals do quite well in estimating the probabilities in 'small' intervals, in which the compound probability does not differ much from the stage-by-stage probability, they fail to appreciate the affect of enlarging the interval i.e., the fact that the probability of success increases. For that reason, for the values of n tested, underestimations increased with n. Another consequence is that subjects fail to fully appreciate increases in the stage-by-stage probability, i.e. the fact that a 'small' increase in the stage-by-stage probability implies a 'large' increase in the overall probability of success.

In the paper we try to get some insight into the relevance of this to 'real-life' problems, such as the equity premium puzzle. It may be that the failure of traditional

[1] A similar game, known as "The gambler's ruin problem", is a classical problem in the random walk literature. Early solutions by Bernoulli and De Moivre are described in Thatcher (1957). For detailed solutions see Ross (1989). This literature was not concerned with the bounded rationality aspects of this problem.

risk measures to explain behavior in many cases is not a case of a bad theory of risk behavior, as much as a simple misjudgment of the objective probabilities by people. See Arrow (1982).

2. Computing the compound probability of success

Let $p_x(t)$ be the probability of getting n after t stages for a player who starts with x. Denote the infinite case by p_x, i.e. $p_x = \lim_{t \to \infty} p_x(t)$. The probabilities p_x satisfy the following system of equations:

$$\begin{aligned} &p_n = 1 \\ &\vdots \\ &p_x = (1-p) * p_{x-1} + p * p_{x+1}, \qquad \text{for } 0 < x < n, \\ &\vdots \\ &p_0 = 0 \end{aligned}$$

Proposition 2.1. The explicit solution of the system, for p ≠ 0.5 is:

$$p(x) = \frac{\left[(1-p)p^{-1}\right]^x - 1}{\left[(1-p)p^{-1}\right]^n - 1}.$$

Proof. This is a system of $n+1$ linear equations in $n+1$ unknowns $(p_0, \ldots, p_n)$. It is easily seen that the determinant of the system is non-zero, hence, the system has at most one solution. Direct verification shows that the equation in the proposition is a solution to the system. Hence it must be the unique solution. □

In Table 1 are the p_x's for a few different x's and p's. $n = 10$.

Table 1
Values of p_x are the probabilities of reaching $n = 10$, starting with x, when p is the stage-by-stage probability

	$p=0.5$	$p=0.55$	$p=0.58$	$p=0.6$	$p=0.65$	$p=0.7$
p_{10}	1	1	1	1	1	1
p_9	0.9	0.96	0.98	0.99	0.998	0.9997
p_8	0.8	0.91	0.96	0.98	0.995	0.9990
p_7	0.7	0.86	0.93	0.96	0.989	0.9976
p_6	0.6	0.79	0.89	0.93	0.978	0.9940
p_5	0.5	0.71	0.83	0.88	0.957	0.9857
p_4	0.4	0.62	0.75	0.82	0.918	0.967
p_3	0.3	0.51	0.65	0.72	0.85	0.92
p_2	0.2	0.37	0.5	0.57	0.71	0.82
p_1	0.1	0.2	0.29	0.34	0.46	0.57
p_0	0	0	0	0	0	0

3. Method and results

In this section, experimental results from three experiments are presented, showing underestimation.

3.1. Experiment 1: Changing the starting amount with experienced subjects

We wanted to check whether having some 'experience' with the game will make the estimates more accurate. First year students in economics at Tilburg University participated in the experiment. The students played a version of the game (see Appendix A) for real money. [2] Each student played privately and independently of the others. In total 28 subjects participated; 10 students started with \$3, 10 with \$5 and 8 students with \$7. The probability used was $p = 0.58$, and $n = 10$. Each session took at most 30 minutes. After playing, subjects were asked indirectly (see Appendix B) about their estimations. This was done in order to check whether the mistakes resulted from confusion created by terms like 'probability' and 'chance'. Their responses are presented in Table 2.

Twenty-four out of the 28 subjects underestimated p_x. This first experiment shows that underestimation exists, even after some experience in playing. An interesting observation is that, on average, playing more stages (more 'experience') did not result in more accurate estimations.

Another experiment, not incentive motivated, used 16 seminar participants (professors and Ph.D students in economics) from Tilburg University as subjects. Each subject was asked to estimate p_x for 4 different x's. Although these subjects are not 'normal

Table 2
The right column in each starting amount gives the estimated p_x for subjects who first played the game for real money. $p = 0.58$ and $n = 10$. For each x, subjects are ordered by their estimation

Subject	Start with \$3	Subject	Start with \$5	Subject	Start with \$7
1	0.85	11	0.95	21	0.85
2	0.75	12	0.80	22	0.83
3	0.70	13	0.70	23	0.80
4	0.62	14	0.70	24	0.75
5	0.58	15	0.58	25	0.58
6	0.58	16	0.58	26	0.58
7	0.58	17	0.58	27	0.58
8	0.40	18	0.58	28	0.44
9	0.25	19	0.37		
10	0.05	20	0.30		
Mean	0.536	Mean	0.614	Mean	0.676
Actual p_x	0.646	Actual p_x	0.834	Actual p_x	0.933

[2] We used Dutch guilders, scaling the game such that the change in prize in every stage was $f\,2.5$ instead of \$1, e.g. when we say that the game started with \$5 it actually started with $f\,12.5$ (at the time $f\,2.5 = \$1.6$). To reduce confusion, we continue presenting the results in dollars.

Table 3
The probabilities of reaching different values of n, when $p = 0.6$, and starting at $n/2$

n	2	4	6	8	10	12	14	16	18	20	50
$p_{n/2}$	0.60	0.69	0.77	0.84	0.88	0.92	0.94	0.96	0.97	0.98	0.99996

people', in the sense that they know more about probability theory than most people, 54 out of the 64 responses underestimated p_x.

3.2. Experiment 2: Changing the size of the interval

Most of the literature on sequential decision problems uses two stage lotteries as a sole representation of dynamics. This is done under the assumption that moving from one-stage to two-stage lotteries captures the essential aspects of dynamics, and moving from two-stage to multi-stage lotteries is trivial. We show that in our random walk example, there is no 'irrationality' in a two-stage set-up, but subjects become 'more irrational' with every stage added.

To do this, different sizes of intervals were used, keeping the rest of the rules the same. In Table 3 are the reduced probabilities for different values of n, when starting with $x = n/2$ (calculated using Proposition 2.1).

As one can see, the reduced probabilities converge to 1 very rapidly. The question raised now is, would subjects, although underestimating the reduced probabilities, understand that they converge to 1?

We used undergraduate students in economics, and gave them monetary incentives to find the accurate probabilities. We had 4 groups of subjects, one with 4 subjects, one with 5, and 2 with 6 subjects each. Appendix C is an example of a questionnaire for $n = 4$. The results are presented in Table 4.

The underestimates are robust even under this treatment (16 out of 21 subjects, but 9 out of 11 for $n = 10$ and $n = 14$), and the size and frequency of underestimations increase with n. Another observation is that the mean of the observations, within the range of $4 \leq n \leq 14$, did not converge to 1, i.e. the expectations are not monotonic and do not differ much from each other.

Table 4
Estimations of $p_{n/2}$ for different values of n

Subject	$n = 4$	Subject	$n = 6$	Subject	$n = 10$	Subject	$n = 14$
1	0.8	1	0.8	1	1	1	0.94
2	0.7	2	0.67	2	0.8	2	0.8
3	0.6	3	0.6	3	0.65	3	0.6
4	0.5	4	0.6	4	0.6	4	0.6
		5	0.6	5	0.35	5	0.6
		6	0.36			6	0.6
Mean	0.65	Mean	0.605	Mean	0.68	Mean	0.69
Actual p_x	0.69	Actual p_x	0.77	Actual p_x	0.88	Actual p_x	0.94

Table 5
Estimations of p_x for different values of p, $x = 5$ and $n = 10$

Subject	Prob 0.55	Subject	Prob 0.6	Subject	Prob 0.65	Subject	Prob 0.7
1	0.9	16	0.99	31	0.9	46	0.99
2	0.68	17	0.8	32	0.85	47	0.83
3	0.59	18	0.77	33	0.82	48	0.82
4	0.59	19	0.73	34	0.8	49	0.7
5	0.57	20	0.7	35	0.74	50	0.7
6	0.55	21	0.69	36	0.7	51	0.7
7	0.55	22	0.65	37	0.65	52	0.7
8	0.55	23	0.64	38	0.65	53	0.7
9	0.55	24	0.6	39	0.65	54	0.7
10	0.55	25	0.6	40	0.65	55	0.65
11	0.55	26	0.6	41	0.6	56	0.6
12	0.55	27	0.52	42	0.58	57	0.58
13	0.55	28	0.5	43	0.5	58	0.3
14	0.45	29	0.2	44	0.5	59	0.17
15	*	30	*	45	0.35	60	*
Mean	0.58	Mean	0.64	Mean	0.66	Mean	0.65
Actual	0.73	Actual	0.88	Actual	0.96	Actual	0.99

* Observations that did not add up to one.

3.3. Experiment 3: Changing the stage-by-stage probability

In this experiment we fixed $x = 5$ and $n = 10$, and varied p. We used 60 first year economics students, in four groups of 15 subjects. Each group was asked about one p. Instructions were similar to those of Experiment 2, but a different reward scheme was used (see Appendix D). The results are presented in Table 5.

Again we see underestimations of p_x (54 out of 57 subjects), and we see that changing p did not change the mean of the estimations which, apart from the case of $p = 0.55$, are almost identical. This implies that subjects were not sensitive to changes in p.

4. Discussion: Anchoring and adjustment heuristic

The evidence indicates that when estimating the compound probability of success (p_x), subjects use the stage-by-stage probability of success (p) as an anchor. Apparently, subjects 'start' with p, anchor to that, and either do not adjust at all, or adjust insufficiently to changes in the parameters. In total, 40 out of the 106 estimations were $p = p_x$. Moreover, if we look at a comparison of the distance $|(p_x - \text{mean})/(p - \text{mean})|$, as done in Table 6, we see that the mean of estimates was at least 2.5 times closer to p than to p_x for all but the $n = 4$ case in Experiment 2.

The consequence of changing the starting amount was tested in Experiment 1, where it was shown that there is some adjustment, always in the correct direction, but the adjustment is insufficient. In Experiment 2 we changed the size of the interval, and

Table 6
Comparison of the distance $|(p_x - \text{mean})/(p - \text{mean})|$ using the mean of estimations, p, and p_x from all the experiments

	p	p_x	mean	$\|(p_x - \text{mean})/(p - \text{mean})\|$
Experiment 1				
1	0.58	0.646	0.536	2.5
2	0.58	0.834	0.614	6.5
3	0.58	0.933	0.676	2.7
Experiment 2				
1	0.6	0.69	0.65	0.08
2	0.6	0.77	0.605	33
3	.6	0.88	0.68	2.5
4	0.6	0.94	0.69	2.8
Experiment 3				
1	0.55	0.73	0.58	5
2	0.6	0.88	0.64	6
3	0.65	0.96	0.66	30
4	0.7	0.99	0.65	6.8

found no sign of adjustment. Experiment 3 shows that estimations are not sensitive to changes in the stage-by-stage probability.

This is not the first attempt to look at estimations of compound probabilities. Bar-Hillel (1973) investigated the hypothesis that the subjective probability of compound events are systematically biased in the direction of their components, resulting in overestimation of the likelihoods of conjunctive events and underestimation of the likelihood of disjunctive events, e.g. a probability of a conjunctive event may be the probability of winning 5 times in a row, and the probability of a disjunctive event is the stage-by-stage probability. Bar-Hillel concluded that "... The probability of the individual stage in a chain of events thus appears to have greater influence on the evaluation of the whole chain's probability than the number of stages in question" (p. 405). This is similar to our result in the sense that people anchor to the probability of the individual stage, and fail to fully appreciate the affect of enlarging the number of stages. However, this work focuses on the probability of a certain path, and not on the probability of outcomes. Note also that increasing the number of stages has an opposite effect as compared to our story, i.e. it reduces the compound probability.

Other studies that report underestimation in multi-stage problems are Wagenaar and Sagaria (1975), and Wagenaar and Timmers (1979). These studies consider a different type of problem, namely estimations of exponential growth. They show that exponential growth is considerably underestimated; people tend to extrapolate exponentially, that is with a constant multiplier for successive steps, but with an exponent that is too small.

5. Application to financial markets

We showed that in a bounded random walk set-up, with positive drift, most subjects underestimate the reduced probabilities of reaching the upper bound. Why should this result interest economists? For example, the traditional finance literature assumes that asset prices in an efficient capital market follow a random walk with positive drift, i.e. that capital markets "have no memory" (Brealey and Myers, 1988, p. 289).[3] Our findings suggest that investors will fail to appreciate the difference in the returns between the short and the long run. For example, a stock that is traded daily, and whose price follows a random walk with a known 'very small' daily positive expected return, may do 'very well' in the long run, much better than people expect it to do on the basis of its daily performance. This implies that investors' perceived risk of that kind of asset is systematically higher than the objective risk and, as a result, assets are undervalued. Another difficulty investors may have is underestimating the chance that a slightly better asset (higher p) will accumulate much larger wealth in the long run.

For example the equity premium puzzle, which is the empirical fact that stocks have outperformed bonds over the last century in a way that is hard to explain with plausible levels of investor risk aversion (Mehra and Prescott, 1985), may partly be the result of investors' misjudgment of risk.

One obvious question is whether markets would 'fix' these underestimations. The problem is that, since the 'objective probabilities' in the stock market are unknown, there cannot be any empirical proof of the kind given in this paper. It may be that the price in a market will reflect the accurate probabilities even if (most) participants are not able to correctly estimate these probabilities. On the other hand, there is evidence that markets are not always efficient (e.g. De Bondt and Thaler, 1994). For an elaborated discussion about the role of risk perception in financial markets see Arrow (1982).

In future research, we would like to address this question, with the aim of tackling Camerer's challenge "Whether judgment and choice violations matter in markets is a question that begs for empirical analysis" (Camerer, 1987, p. 981).

Acknowledgements

I am grateful to Eric van Damme for guiding me in my work, to the referees, and to all the people at CentER who helped in running the experiment. All mistakes are mine.

Appendix A

A.1. Instructions for subjects who played the game for real money

In a few minutes we will give you \$5 and ask if you want to participate in a game in which at every step you can either win or lose \$1. The chance of winning \$1 is 58% and the chance of losing \$1 is 42%. If you choose to leave the game, you can stop and take the \$5. If you decide to play, you will either have \$4 or \$6 after the first stage. Then,

[3] This approach is controversial nowadays (e.g. Fama, 1991, or De Bondt and Thaler, 1994). Yet as a first approximation it is still accepted, and that is enough for our case.

you can leave the game with your money, or participate in the next stage, in which, again, you can either win or lose $1 with the same chances as before. The game goes on until either you choose to stop, or your money reaches {0 or $10}, or after 100 stages. [4]

Appendix B

B.1. Indirect method for finding estimations (note that the question is phrased such that it is equivalent to the initial problem)

Say that we take 100 students and let them play the game, with one difference: they will have to play till $0 or $10. Can you guess how many of them will end up with $0 and how many with $10?

Appendix C

C.1. Estimations for different sizes of intervals with monetary incentives (for n = 4)

Please answer the following question, which is also given to other students in the room. After all of you have finished answering, we will collect and check the answers. We will find the best answer and give $10 to the student who gave it. If more than one student gives the best answer, we will split the money between the students who gave this answer.

The game: Mr. X is given $2 and then a series of lotteries take place. In each lottery he either wins or loses $1. The chance of winning $1 is 60% and the chance of losing $1 is 40%. So, after the first lottery, Mr. X will either have $3 (with 60% chance) or $1 (with 40% chance), and so on. The lotteries will be conducted till Mr. X will either have $0 or $4.

The problem: What do you think is the chance that Mr. X will:
(a) finish with $4? ____
(b) finish with $0? ____

Appendix D

D.1. You will be paid according to the following rule

You will start with $15, and for every 1% of 'mistake' $1 will be deduced from your payoff. The mistake is the absolute value of [your guess (in percentages) minus the

[4] The probability that the game will not end within 100 stages is less than 0.001, hence p_x is relevant even for $t = 100$. For a discussion of the use of this restriction see Gneezy (1995).

actual chance]. For example, if you guess accurately, you get $15. If you make a 10% mistake (either overestimate or underestimate), you get $5. If your mistake is bigger or equal to 15% you will not be paid at all.

References

Arrow, K.J., 1982. Risk perception in psychology and economics. Economic Inquiry 20, 1–9.
Bar-Hillel, M., 1973. On the subjective probability of compound events. Organizational Behavior and Human Performance 9, 396–406.
Brealey, R. and S. Myers, 1988. Principles of corporate finance (3rd ed.). New York: McGraw-Hill.
Camerer, C., 1987. Do biases in probability judgment matter in markets? Experimental Evidence. American Economic Review, December, 981–997.
Camerer, C., 1995. 'Individual decision making'. In: J.H. Kagel and A.E. Roth (Eds.), Handbook of experimental economics. Princeton, NJ: Princton University Press.
De Bondt, W. and R.H. Thaler, 1994. 'Financial decision-making in markets and firms: A behavioral perspective'. In: R. Jarrow, V. Maksimovic, and W.T. Ziemba (Eds.), Amsterdam: North-Holland.
Fama, E.F., 1991. Efficient capital markets: II. Journal of Finance 46, 1575–1617.
Gneezy, U., 1995. Updating the reference level: Experimental evidence. Draft.
Hammond, P.J., 1988. Consequentialist foundations for expected utility. Theory and Decision 25, 25–78.
Kahneman, D., P. Slovic and A. Tversky, 1982. Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press.
Kahneman, D. and A. Tversky, 1979. Prospect theory: An analysis of decision under risk. Econometrica 47, 263–291.
Machina, M.J., 1989. Dynamic consistency and non-expected utility models of choice under uncertainty. Journal of Economic Literature 27, 1622–1668.
Mehra, R. and E.C. Prescott, 1985. The equity premium; A puzzle. Journal of Monetary Economics 15, 145–161.
Ross, S.M., 1989. Introduction to probability models. London: Academic Press.
Savage, L.J., 1954. The foundation of statistics. New York: Wiley.
Thatcher, A.R., 1957. Studies in the history of probability and statistics. VI. A note on the early solutions of the duration of the play. Biometrika 44, 515–518.
Wagenaar, W.A. and S.D. Sagaria, 1975. Misperception of experimental growth. Perception and Psychophysics 18, 422–426.
Wagenaar, W.A. and H. Timmers, 1979. The Pond-and-Duckweed problem. Acta Psychologica 43, 239–251.