MAGYAR TUDOMÁNYOS AKADÉMIA
Közgazdaság- és Regionális Tudományi Kutatóközpont

MTA KRTK KTI

Centre for Economic and Regional Studies
HUNGARIAN ACADEMY OF SCIENCES

MŰHELYTANULMÁNYOK DISCUSSION PAPERS

MT-DP – 2013/42

Gambler's fallacy in the classroom?

HUBERT JÁNOS KISS - ADRIENN SELEI

INSTITUTE OF ECONOMICS, CENTRE FOR ECONOMIC AND REGIONAL STUDIES,
HUNGARIAN ACADEMY OF SCIENCES BUDAPEST, 2013

Discussion papers
MT-DP – 2013/42

Institute of Economics, Centre for Economic and Regional Studies,
Hungarian Academy of Sciences

Gambler's fallacy in the classroom?

Authors:

Hubert János Kiss
research fellow
in the Momentum(LD-004/2010) Game Theory Research Group
Institute of Economics,
Centre for Economic and Regional Studies
Hungarian Academy of Sciences and
Eötvös Loránd University - Department of Economics
and Affiliate Fellow at CERGE-EI, Prague
email: kiss.hubert.janos@krtk.mta.hu

Adrienn Selei
research associate
Regional Centre for Energy Policy Research
and Eötvös Loránd University
email: adrienn.selei@uni-corvinus.hu

December 2013

ISBN 978-615-5447-03-7
ISSN 1785 377X

Gambler's fallacy in the classroom?

Hubert János Kiss - Adrienn Selei

Abstract

Does students' hand tremble after marking three consecutive identical answers in a multiple choice test? We design an experiment to study if the likelihood to change incorrectly to a different answer than the last one depends on the number of identical previous answers. We do not find a clear treatment effect, but observe that indeed the likelihood to change to an incorrect answer increases in the number of identical previous answers given by the student, even after controlling for how prepared (s)he was overall and how certain (s)he was that the answer to a given multiple choice question is correct. We claim that this behavior possibly is a reasonable reaction to previous exam experience.

Keywords: belief, experiment, gambler's fallacy, multiple choice

JEL Classiffication: C93; D03; D84

Acknowledgement:

Financial support from the Spanish Ministry of Science and Innovation under the project ECO 2011-25349 and from the Hungarian Scientific Research Fund (OTKA) under the project PD 105934 are kindly acknowledged.

A szerencsejátékos tévedése az osztályteremben?

Kiss Hubert János – Selei Adrienn

Összefoglaló

Megremeg a diákok keze, miután három ugyanolyan választ adtak feleletválasztós tesztben? Egy kísérletet végzünk, amelyben megvizsgáljuk, hogy miként függ a korábbi azonos válaszok számától annak valószínűsége, hogy egy diák helytelenül a következő kérdésre más választ ad. Nem találunk világos kezelési hatást, de az látható, hogy a helytelen válaszra való áttérés valószínűsége nő a korábbi azonos válaszok számában, még akkor is, ha figyelembe vesszük a diákok felkészültségét és azt, hogy mennyire voltak biztosak a válasz helyességében. Úgy gondoljuk, hogy ezen döntéshozatal ésszerű reakciónak tekinthető a korábbi vizsgatapasztalatok fényében.

Tárgyszavak: vélekedés, kísérlet, a szerencsejátékos tévedése, feleletválasztós teszt

JEL kódok: C93; D03; D84

Gambler's fallacy in the classroom?

Hubert János Kiss*
Eötvös Loránd University

Adrienn Selei
Eötvös Loránd University

December 4, 2013

Abstract

Does students' hand tremble after marking three consecutive identical answers in a multiple choice test? We design an experiment to study if the likelihood to change incorrectly to a different answer than the last one depends on the number of identical previous answers. We do not find a clear treatment effect, but observe that indeed the likelihood to change to an incorrect answer increases in the number of identical previous answers given by the student, even after controlling for how prepared (s)he was overall and how certain (s)he was that the answer to a given multiple choice question is correct. We claim that this behavior possibly is a reasonable reaction to previous exam experience.

Keywords: belief, experiment, gambler's fallacy, multiple choice
JEL Classification: C93; D03; D84

1 Introduction

Beliefs play an important role in economic theory. Standard models assume that people's view on the distribution of the states of the world is correct on average. However, there is a growing body of evidence that documents deviations from correct beliefs and shows the ways beliefs may be systematically inaccurate (see DellaVigna, 2009 for a review). A deviation from standard belief formation is often based on the inadequate weighting of recent available information. Concretely, consider the case when people are making the same kind of choice repeatedly. When observing a sequence of the same outcome they may have one of the three following belief: (1) that the streak is irrelevant, (2) that the streak will continue, or (3) that the streak will stop (Burns and Corpus, 2004). An example of the belief that a streak should continue is the hot hand fallacy (Gilovich et al., 1985), whereas the gambler's fallacy (see for instance Ayton

*Financial support from the Spanish Ministry of Science and Innovation under the project ECO 2011-25349 and from the Hungarian Scientific Research Fund (OTKA) under the project PD 105934 are kindly acknowledged. Eötvös Loránd University - Department of Economics; Research fellow in the "Momentum" (LD-004/2010) Game Theory Research Group Institute of Economics Centre for Economic and Regional Studies of the Hungarian Academy of Sciences; and Affiliate Fellow at CERGE-EI, Prague. Email: kiss.hubert.janos@krtk.mta.hu.

and Fischer, 2004) is an illustration of the tendency to believe that a streak of events is likely to end.

This paper studies if students when taking a multiple choice test make some inference about the right answer based on the previous pattern of answers given. We are interested also in whether the conclusions that they draw (if it is the case) are incorrect or not. That is, after marking three consecutive B's do students believe that the next answer is less likely to be another B? And if it is the case, are they wrong? To test if such a phenomenon at first glance similar to the gambler's fallacy may be at play during exams we carried out a field experiment. In several exams of the same academic subject we divided students in two groups randomly and they received the same test (25 multiple choice questions in the same order and 5 short essay questions) with only one difference. In the control group the right answers varied often between questions, while in the treatment group answers were arranged in a way that sequences of identical answers were frequent. We controlled also for how certain students were in their answers in an incentivized way and also for how prepared the students were through the essay questions. The results show that there is no clear treatment effect, but the longer the sequence of identical previous answers, the more probable is that a student incorrectly marks a different answer. While this behavior is reminiscent of the gambler's fallacy, we argue that it is more likely that students' decision is driven by their previous experience. Multiple choice tests generally exhibit an alternating pattern of correct answers and when uncertain about the right answer students tend to mark answers so that this varying pattern is reproduced.

2 Experimental Design and Conjectures

2.1 Design

There were 5 sessions, corresponding to the exam dates in May and June, 2012. In total 153 students (105 women) participated in the experiment.[1] Subjects were undergraduate students enrolled in the International Studies Program of the Faculty of Social Sciences on the Eötvös Loránd University (Budapest, Hungary). Students could freely choose when to take the exam on International Economics. There was one exam each week during the exam period. The questions varied from exam to exam, though there were questions which appeared in several exams.

Each session had two treatments. In both treatments, the students had to answer 25 multiple choice questions and 5 short essay questions. Multiple choice questions were valued 2 points if the answer was correct and zero otherwise. There was no point deduction for incorrect answers. In each essay question students received 10 points for the right explanation and less if important details

[1] 126 students took the exam once. Those who failed or wanted to improve their mark could take the exam again. There were 25 students who took the exam twice and 2 students did it three times.

were left out.

Questions in both treatments were the same and in the same order, there was only one difference. In treatment 0 (our control), the right answers varied frequently, there were 23 alternations between answers. In treatment 1 (where the possible answers were the same as in treatment 0), there were a lot less alternations (only 9), streaks of the same answer were abundant. These were the pattern of answers in the treatments:

Question	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Tr 0	B	A	D	A	D	B	C	D	A	B	D	D	B	A	D	C	A	B	C	D	B	A	C	D	B
Tr 1	A	B	B	B	D	D	D	D	A	A	A	C	C	C	C	D	A	B	B	B	B	A	D	D	D

Figure 1. The pattern of correct answers in the two treatments

From session to session we changed the pattern of answers slightly by renaming the answers: for instance, the A answers in session 1 became B answers in session 2. In this way the structure of the answers and the number of alternations remained the same across all sessions.

When the students arrived, they could freely choose a seat. Before starting the exam we rearranged a bit the seating so that students be equally seated, with sufficient space between them in order to minimize the possibility of copying. At the beginning of the exam we explained loudly in front of the students the rules of the exam. Once the exam started talking was forbidden and we suspended those students who were caught either talking, copying from other or using illicit materials (it was a closed book exam). We monitored intensely throughout the exam to minimize cheating. Exam sheets were handed out in a way that students next to each other received different versions (treatment 0 and treatment 1).

Note that there is a potential confounding if we do not control for how certain students were about their answers. Imagine that a student after two B answers changes to C, B still being the right answer. Without knowing how certain (s)he was we may classify it as a change to an incorrect answer due to the students' trembling hand. However, if the certainty point reveals that (s)he really thought that C was the right answer, then we can exclude this case as a manifestation of the trembling-hand hypothesis. Thus, we told the students when explaining the rules of the exam that they could earn 3 extra points if they assessed correctly their performance. To do so they were asked to rate after each multiple choice question how certain they were if their answer was correct. These certainty points ranged from 1 to 5, 5 (1) meaning that somebody was totally (un)certain about the correct answer. In the instructions (for details see the Appendix) we specified that if the number of right answers divided by 5 is within a certain range (± 0.25) of the mean of the certainty points, then the student's self-assessment was considered accurate and (s)he received 3 extra points. For instance, a student that had 20 correct answers received the extra points, if the average of her certainty points lied between 3.75 and 4.25.[2]

[2] Unfortunately, this is not a perfect control since it has an inherent upward bias in the case of the most uncertain answers. When a student is totally uncertain about the right answer,

These potential extra points served to incentivize subjects to think hard over the certainty points. Regarding the certainty points, we emphasized two issues at the beginning of the exam. First, we stressed that filling in the certainty points is not compulsory. Second, we assured students that certainty points are used for research purposes (we did not specify the nature of the research) and do not influence grades negatively. More precisely, we made clear that if somebody is guessing (which can be revealed through low certainty points) but happens to give the right answers, (s)he will get the corresponding points and grade, no points are deducted for guessing. Note that there was no value of strategic signaling by giving high certainty points to suggest that somebody is sure in the answers and knows the material well. High certainty points did not affect the grade *per se*, only if they corresponded with the actual points earned.

2.2 Conjectures

Our first conjecture posits the expected existence of a treatment effect. If students tremble when observing streaks of previous answers, then they are more likely to score less in treatment 1.

Conjecture 1 *All else being equal, on average students in treatment 1 score less points in the multiple choice questions than students in treatment 0.*

The second conjecture is about the likelihood of changing incorrectly the answer from one question to another. Hence, we focus on those cases in which this change is not justified, that is a student marks a different answer than the previous one even though the right answer is the same. We conjecture that the more identical previous answers marked by the student, the more likely is that such an incorrect change occurs.[3]

Conjecture 2 *All else being equal, the probability of an incorrect change in answers is increasing in the number of identical previous answers.*

We remark that there are many studies about behavior when taking a multiple choice test. These analyze various topics like guessing (e.g. Burgos, 2004; Espinosa and Gardeazabal, 2010) or gender differences in multiple choice test (for instance, Siegfried, 1979; Ballard and Johnson, 2005; Marín and Rosa-García, 2011). To the best of our knowledge, no study focuses on the topic that we are investigating.

---

then we would like her/him to give 1 certainty point, but by pure guessing the student has 25% chance of hitting the right answer, so s(he) may find optimal to give a higher point. In spite of this imperfection we hope to have captured at least partially how sure students were in their answers.

[3] Note that what matters is the number of identical previous answers that the student thought right and marked on the answer sheet, and not the ones that are actually right.

3 Results

First, we present some descriptive statistics about the points received in the exam. Since in session 2 and 5 just a small number of students (17 and 9) took part we executed detailed analysis on the rest of the sessions. However, we considered the observations obtained in these sessions in the overall analysis Table 1 shows the mean points achieved in multiple choice test, the average of certainty points and the average of essay points according to treatments.

	Points in multiple choice test		Average of certainty points		Average of essay points	
	Treatment 0	Treatment 1	Treatment 0	Treatment 1	Treatment 0	Treatment 1
Date 1	35.5 (30)	32.64 (28)	3.81 (30)	3.78 (28)	25.67 (30)	27.68 (28)
Date 3	34.42 (19)	34 (20)	3.94 (16)	3.98 (13)	34.42 (19)	33.45 (20)
Date 4	34.67 (24)	33.94 (35)	3.97 (23)	3.99 (27)	33.54 (24)	28 (35)
Overall	34.66 (86)	33.83 (96)	3.86 (79)	3.92 (81)	29.73 (86)	29.67 (96)

Number of observations in brackets. The number of observations for the certainty points may be lower since not all students filled in the certainty points.

Table 1. Multiple choice test, certainty and essay points

Using both Wilcoxon ranksum test and t-test w efind that there is no significant difference in certainty and essay points, except for the essay points at date 4. There in treatment 0 the essay points were significantly higher than in treatment 1 at 10 % significance level. The number of observations indicates that not each student filled in the certainty points as it was not compulsory. Those who filled in the certainty points achieved better results in multiple choice test (34.59 vs. 31.55) and the difference is significant at 5% significance level according to the one-sided t-test (H1: points of those who filled in > points of those who did not), though the Wilcoxon ranksum test fails to reject the hypothesis that points in both groups were the same. Henceforth, we restrict our attention to those who marked how certain they were regarding their multiple choices.

Since differences are generally not significant regarding the certainty and the essay points, so differences in the multiple choice test are unlikely to be driven by differences in the student pool that was assigned to the treatments. Although multiple choice test points are higher in treatment 0 in all cases, t-tests and the non-parametric Wilcoxon ranksum test fail to reject the null hypothesis that points in treatment 0 and 1 are the same, both for multiple choice test and certainty points. The only exception is the multiple choice test point at date 1 for which the one-sided t-test (H1: points in treatment 0 > points in treatment 1) reveals significant difference at 10 % significance level. These tests suggest that the treatment did not produce a significant effect.

To test the first conjecture we run an OLS regression where the dependent variable is how many points a student scored in the multiple choice test. The independent variables include: i) a treatment dummy (0 for normal test and

1 for the treated test); ii) certainty (the average of the certainty points which reflects how sure students were about the correct answers; it ranges from 1 to 5); iii) essay (the points obtained for the essay questions as a proxy for how well prepared were the students for the exam; it ranges from 0 to 50); iv) female (1 for women and 0 for men) and v) assessment (a dummy that is 1 if the average of certainty points was within a predetermined interval relative to the average of the points scored in the multiple choice test, reflecting how well students assess their own performance).

We run the above regression for each date with a sufficient number of students (date 1, 3 and 4) and for the pooled data. For the last case we include date dummies (for each exam date, and date 1 being the baseline) and standard error is clustered on the individual level. The results are as follows:

Dependent variable: Points in multiple choice test				
	date 1	date 3	date 4	overall
treatment	-3.03** (1.18)	-0.20 (2.19)	2.13 (1.51)	-0.77 (0.79)
certainty	3.54*** (1.27)	2.53 (2.00)	4.18*** (1.48)	4.12*** (1.21)
essay	0.22*** (0.06)	0.19 (0.14)	0.30*** (0.07)	0.23*** (0.06)
female	-0.55 (1.23)	-3.81 (2.60)	-0.02 (1.54)	-0.64 (0.88)
assessment	3.83*** (1.28)	5.32* (2.64)	4.95*** (1.62)	4.74*** (0.77)
constant	15.20*** (4.52)	19.32** (7.44)	6.4 (5.4)	11.57*** (3.91)
N	58	29	50	160
Prob>F	0.00	0.02	0.00	0.00
R-squared	0.6	0.43	0.6	0.52

*/**/*** denote 10%/5%/1% significance level. Standard erros in brackets.

Table 2. OLS regression results

Treatment has a significant effect at date 1. Points obtained in the multiple choice test are 3 points lower in the treated group even if we control for the other variables. However, at the other dates and when looking at the overall data treatment has no significant effect. This result suggests that the treatment effect is not pervasive as already foreshadowed by the tests. Even if we manipulate extensively the alternation rates across treatments, it does not cause a significant difference in the achieved multiple choice test points.

It is worth noting that, all other things held constant, certainty has almost in each case a highly significant positive effect, indicating that students who are more certain about the correct answers achieve more points. We find also that students who are able to evaluate their performance well generally achieve 4-5 points more than those who do not. Unsurprisingly, students who are better prepared (hence have higher essay points) also score more points in multiple choice test. We do not report in the last column of the table the date effect since date dummies are not significant.[4]

The second conjecture suggests that the likelihood of incorrect decisions is increasing in the number of identical answers. To test this conjecture, we examine the determinants of the likelihood of incorrect changes. We run a probit regression in which the dependent variable IC is a dummy with value 1 if there was an incorrect change. The variable assessment, essay and female are defined as before. The variable *same* equals the number of identical answers given by the student (not the actually right answers) for the preceding multiple choice questions. Certainty_q is the certainty point assigned to a given question (and not the average of the certainty points!). We also include date dummies for each exam day, the first exam day being the baseline. We cluster standard errors on individual level.

Since in treatment 0 the alternation rate is very high, so there incorrect changes are very rare by construction. Therefore, we limit ourselves to the data coming from treatment 1. Table 3 shows the marginal effects after our probit regression which are evaluated at the means of the explanatory variables.

Dependent variable: Probability of incorrect change				
	date 1	date 3	date 4	overall
same	0.037**	0.036*	0.017	0.020***
certainty_q	-0.052***	-0.057***	-0.041***	-0.051***
essay	-0.002**	0.001	-0.002**	-0.002**
assessment	-0.043**	-0.039	-0.040	-0.043**
female	0.004	0.059***	-0.003	0.005
N	685	284	675	1969
Prob>chi2	0.00	0.00	0.00	0.00

*/**/*** denote 10%/5%/1% significance level.

Table 3. Marginal effects of the probit regression

The probit analysis supports Conjecture 2. The number of identical previous answers (the length of the streak before the examined question) represented by the variable *same* has a significant positive effect in all but one session and when considering the pooled data the effect is significant at 1 % significance

[4] If we do not restrict our attention to those who filled in the certainty points, but consider the whole sample, then still there is a treatment effect at date 1 but not at the other date and when considering the pooled data. Essay is highly significant at all dates and also when taking all observations together. When looking at the pooled data, the date dummy is significant only at date 3, indicating that at that exam students scored less points in the multiple choice test relative to date 1.

level. If the streak before a given question contains one more identical answer it increases the probability of incorrect change by 2-4%, all else being the same. The certainty point of the question has a significant negative effect on the probability of possible changes. One point increase in the certainty point decreases the probability of incorrect change by 4-5 %, ceteris paribus. The essay points as a proxy of how prepared the student was show a significant effect in all but one case: ten additional points achieved by the student decrease the probability of incorrect change by 1-2 %, all else held constant. Assessment has a negative effect and is significant overall, suggesting that students who evaluate better their performance commit less incorrect changes. The variable female is highly significant in one session, but overall its effect does not seem to be important. From the last column we omitted the date dummies since they are not significant except the dummy for date 4 (at 10 % significance level).

4 Discussion

Although we do not find a clear treatment effect, it seems that there are forces at work that make that students tend to mark a different answer after a sequence of identical answers. Most notably, the more identical previous answers, the more probable that a student commits an incorrect change. Can we rationalize this finding? In the case of the gambler's fallacy the underlying data generating process is entirely random, subsequent events are independent. In our case, the professor generates the answer pattern, so the students' tendency to alternate the answers when uncertain about the right answer may just express their expectations about the pattern chosen by the professor. Students may be used to multiple choice tests in which the correct answers tended to vary frequently. To check if this explanation has some bite we gathered multiple choice tests from our colleagues to see if answers are serially correlated or not. Having checked 20 exams from last years we found that indeed the pattern of right answers is highly variable. We measured the serial correlation as 1 minus the number of alternations between correct answers divided by the number of possible alternations.[5] This measure varies between 5 and 29 % (20 % on average) in the case of the tests we examined, while in our treatment 1 it is 62.5 %. Hence, possibly students that are unsure about the right answer use their experience with multiple choice tests as a guide and mark a different answer when observing a streak of previous identical answers. Without knowing anything about the pattern-setting behavior of the professor this strategy seems reasonable. Therefore, possibly there is no fallacy, but a potential reaction to previous experience.

[5] For example if there were 22 multiple choice questions in an exam with 17 alternations between answers (the correct answers were similar to the previous ones 4 times), it yields the value 1-$\frac{17}{21}$, that is 19 percent.

5 Acknowledgements

We thank András Kiss and participants of the the Annual Conference of the Hungarian Economic Association 2012 for the helpful comments. Kiss kindly acknowledges financial support from the Spanish Ministry of Science and Innovation under the projects ECO2011-25349 and from the Hungarian Scientific Research Fund (OTKA) under the project PD 105934.

References

[1] Ayton, P. and Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: two faces of subjective randomness? Memory and Cognition, 32(8). 1369-1378.

[2] Ballard, C. and M. Johnson (2005). "Gender, expectations, and grades in introductory microeconomics at a US university", Feminist Economics, 11(1). 95-122.

[3] Burgos, A. (2004). Guessing and gambling. Economics Bulletin 4(4). 1-10.

[4] Burns, B. D. and Corpus, B. (2004). Randomness and inductions from streaks: "Gambler's fallacy" versus "hot hand". Psychonomic Bulletin & Review, 11. 179–184.

[5] DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. Journal of Economic Literature, 47(2). 315–72.

[6] Espinosa, M. P., and Gardeazabal, J. (2010). Optimal correction for guessing in multiple-choice tests. Journal of Mathematical Psychology, 54(5), 415-425.

[7] Gilovich, T., Vallone, R. and Tversky, A. (1985) The Hot Hand in Basketball: On the Misperception of Random Sequences. Cognitive Psychology, 17. 295–314.

[8] Marín, C. and Rosa-García, A. (2011): Gender bias in risk aversion: evidence from multiple choice exams. mimeo

[9] Siegfried, J. (1979). Male-Female differences in economic education: a survey. Journal of Economic Education, 10, 1-11.

6 Appendix

International Economics 2. - Group A

Date: 29/06/2012.

You have 90 minutes to complete the exam.

The exam has two parts. Part one consists of a multiple choice test with 25 questions. There is only one right answer to each question. The correct answer scores 2 points, incorrect or unanswered questions yield zero point. Mark the answers with an X in the table below. You are not allowed to change your answer in the table, corrected answers give zero point! When evaluating the exam I only consider the answers marked in the table. In part two of the exam you find five essay questions, each worth 10 points.

Grades are determined as follows:

0-49 points – fail

50-59 points – pass

60-69 points – satisfactory

70-79 points – good

80-100 points – outstanding

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A																									
B																									
C																									
D																									

In the last row of the table you may mark on a scale of 1 to 5 how sure you are about the answer given (1: totally unsure, ..., 5: totally sure) This last row does not affect grading: if somebody is unsure about the answer, but marks the right answer, then she / he will get the 2 points. However, if these "certainty points" correspond to the actual points, then 3 extra points are awarded. More precisely, if the (number of good answers/5) indicator is within ± 0.25 of the average of the certainty points, then those extra points are given. For example, if somebody obtains 20 good answers and the the average of the certainty points is between 3.75 and 4.25, the she / he receives 3 extra points.

Filling in this last row is NOT compulsory. We would like to use these data for research purposes.