University students reflecting on a problem involving uncertainty: what if the coin is not fair?

Francesco Beccuti, Ornella Robutti

University students reflecting on a problem involving uncertainty: what if the coin is not fair?

Francesco Beccuti[1] and Ornella Robutti[2]

[1,2] University of Turin, Italy; francesco.beccuti@unito.it, ornella.robutti@unito.it

How do university master's students in mathematics reason about a simple problem involving coin tosses? We answer to this question by analyzing the participants' written reflections in terms of recency and equiprobability effects understood non-normatively. Overall, the participants display a strong tendency to deem the problem's coin to be fair despite indication that this should be the case and contrary to explicit evidence presented in the problem. In view of this, we further elaborate on the connection between the participants' answers and their education in mathematical probability.

Keywords: Probabilistic reasoning, university students, recency effects, equiprobability effect.

Introduction

The recency and equiprobability effects are well-known phenomena related to people's interpretations of problems involving uncertainty. The *positive recency effect* is the tendency to interpret the manifestation of some event as evidence that the same event is likely to happen again in the future. Conversely, the *negative recency effect* is the tendency to interpret the manifestation of some event as evidence that the same event is less likely to happen again in the future. Moreover, the *equiprobability effect* is the tendency to judge a set of events as all equally likely to happen.

These definitions differ from the ones usually employed in the literature (e.g., Gauvrit & Morsanyi, 2014; Morsanyi & Szucs, 2014; Chiesi & Primi, 2009) essentially in the fact that they are explicitly non-normative: i.e., they do not account for whether the aforementioned effects are against the assumed normatively-correct interpretation of the problem under consideration. Indeed, these effects have been often termed "biases" or "fallacies" or "misconceptions" as they have been mostly problematized with reference to cases in which the respondents' answers to word-problems were considered wrong (cf. Chernoff & Sriraman, 2020; Batanero, 2020).

For instance, an influential article by Fischbein and Schnarch (1997) contained the following question posed to, among others, 18 prospective teachers specializing in mathematics:

> When tossing a coin, there are two possible outcomes: either heads or tails. Ronni flipped a coin three times and in all cases heads came up. Ronni intends to flip the coin again. What is the chance of getting heads the fourth time? (p. 98)

The answer "equal to the chance of getting tails" was taken to be the correct answer (given by 17 students). The answer "smaller than the chance of getting tails" was taken to be evidence of a negative recency effect (given by no student), while the answer "greater than the chance of getting tails" was taken to be evidence of a positive recency effect (given by one student).

As to this student, was he or she simply incorrect or perhaps just interpreting the described fictional situation differently? After all, nothing in the statement of Fischbein and Schnarch's word problem as reported suggests that the hypothetical coin tossed by Ronni has to be considered a *fair* coin or that the way in which Ronni tosses the coin is not biased towards heads. As Gigerenzer (1991, 1996) has argued, probabilistic word-problems usually do not have only one correct answer over which there exists unquestioned consensus. It is true that often people's answers deviate problematically from the generally accepted norm. However, this discrepancy could be caused by the respondents' divergent interpretation of the situation presented to them (Chiesi & Primi, 2009, p. 152). How come that, we may in turn ask, the great majority of the students tested by Fischbein and Schnarch were keen to interpret the problem as involving a fair coin and a fair toss despite the information provided? According to the non-normative definition given above, we may also read these responses as manifestations of an equiprobability effect which we may also problematize. And what about the other responses? Unfortunately, Fischbein and Schnarch did not report on the participants' reflections and hence it is not possible to reconstruct their reasoning.

In this paper we address the following research question by means of a problem involving coin tosses analogous to the one employed Fischbein and Schnarch: *how do university students in mathematics reason about uncertainty?* We answer by means of an analysis of the students' written responses based on the definitions of recency and equiprobability effects given above. Furthermore, nuancing the students' reflections will give us the opportunity of discussing them in relation to their usage of acquired probabilistic concepts and ideas. Notice that our intention in this study is mainly descriptive: while we will discuss the problematic nature of many of the participants' responses, we will not attempt here to suggest any related pedagogical or curricular ameliorations.

Summary of relevant research

The study of Fischbein and Schnarch tested primary and secondary school students together with, as said, prospective teachers specializing in mathematics. Concerning the problem quoted above, the authors found that the negative recency effect decreases with age, while the positive recency effect is almost negligible. Normative equiprobability answers to this problem also increase with age. The authors hypothesized that their findings could be linked to the participants' education in probability. Rubel (2007) in turn tested various problems involving coin tosses on secondary students and analysed the justifications the participants gave for their responses (classified according to a belief framework). Overall, she showed that older children are not more subject to errors connected equiprobability and recency than younger children. However, she hypothesized that this could be due to the fact that the students she tested had only limited exposure to instruction in probability.

Chiesi and Primi (2009) tested primary school children and university students with problems involving drawing marbles with replacement from two bags. As to problems involving bags containing an equal number of marbles, they showed that the positive recency effect decreases with age while the negative recency effect increases and is found at remarkable rates in university students. The equiprobability effect linked to such questions also increases with age but is stable at a significant rate after Grade 5. When considering bags with different numbers of marbles, non-normative equiprobability answers were also remarkably chosen by university students. Furthermore, Morsanyi et al. (2009) enacted a cross-educational and cross-national study testing mainly university students in psychology with various problems involving uncertainty. They found that non-normative equiprobability answers are correlated to the participants' formal education in statistics. Overall, researchers have established an interesting link between the equiprobability effect (understood normatively as a bias) and education: the effect increases with age and is correlated to education in probability and statistics (Chiesi & Primi, 2009; Morsanyi et al., 2009; Saenen et al., 2015). In particular, Chiesi and Primi (2009) and Morsanyi et al. (2009) explicitly argued that the equiprobability effect could be a consequence or a "side-effect" of formal education.

These findings seem to echo insights from qualitative socio-cultural research focusing on word problems. These have suggested that many people approach word problems as activities reduced to the execution of some predetermined operations or algorithms without consideration of possible reality-constraints that might be implied by the problem itself and often neglecting their own everyday knowledge (cf. Verschaffel et al., 2020). Verschaffel and colleagues – surveying a large set of empirical studies mostly involving lower grades pupils – conclude that such conducts develop as a consequence of education in schools, which tacitly but systematically determines how students have to behave. Comparable conducts were reported by research on theorem proving testing both compulsory school pupils (e.g., Harel & Sowder, 1998; Paola & Robutti, 2001) as well as undergraduate students in mathematics (Stylianou et al., 2006): students seem to privilege justifications of mathematical theorems via ritualistic and/or authoritarian proof schemes, possibly as an effect of their years-long exposure to traditional teaching practices.

The present study

Participants

The participants are 84 students (34 males and 50 females of median age 24) of the course "Didactics of Mathematics 1" within the master's degree in mathematics at the University of Turin, Italy. This is a program focusing on advanced mathematics. Access to the program is conditional on holding a bachelor's degree in mathematics obtained with good proficiency from a recognized university. All the participants have passed at least one compulsory university course in probability and statistics (typically presented within an axiomatic framework). The following experiment was performed by testing two cohorts of students in two subsequent years with the same procedure involving a short computer-based questionnaire.

Procedure

The questionnaire was divided into two tasks requiring the students to work individually. The questionnaire was part of a larger written assignment which included a variety of mathematical questions and problems proposed to the students during the class as a test aiming to evaluate their general competence in solving mathematical problems. As to the questionnaire relevant for this study, each participant was presented with the following multiple-choice question:

Task 1

Sara tosses for ten times a coin and for ten times she obtains heads. Sara then asks to Piero to bet on the outcome of the next toss. Piero then bets five euros on the next toss resulting in heads again.
Do you agree with Piero's choice?

☐ Yes ☐ No ☐ In part ☐ I am not sure

After the student submitted the answer, the computer immediately presented a related second task:

Task 2

Explain your reasoning.

The student could answer by submitting a text possibly containing mathematical symbols. The answers were automatically recorded by the computer.

Explanation of the choices

The possibilities provided in the multiple-choice question of Task 1 were selected as to allow the participants to nuance broadly their judgement concerning the fictional situation presented. We chose to proceed in this way in order to not necessarily force on the participant a yes/no answer, but rather to stimulate reflection over the problem in view of Task 2, the main focus of this paper. We leave a detailed analysis of the interrelations of the participants' answers to Task 1 and Task 2 to a subsequent paper. Task 2 in turn was formulated as an open question in order to offer to the participants opportunity for ample reflection in consideration of our research question. Overall, the problem was formulated as asking for a judgement towards a fictional decision-problem rather than as a problem of direct estimation of probability, likehood, or chance. This choice was made in order to induce in the participants a detached viewpoint towards the situation described and in order to avoid the difficulties of interpretation connected with most-likely/least-likely questions (cf. Rubel, 2007, pp. 533–534). Nonetheless, we hypothesized that the students would themselves refer to the possibility of estimating the outcome of a hypothetical 11^{th} coin toss. Thus, we anticipated to be able to nuance the participants' reflections on the problem in relation to recency and equiprobability and in connection to acquired probabilistic concepts.

Method of analysis

For the present article we concentrate primarily on the analysis of the written responses to Task 2. Concerning the latter, we classified each answer according to a deductive coding procedure (cf. Braun & Clarke, 2006, pp. 83-84) based on the definitions of equiprobability and recency effects given in the introduction. Having found that the great majority of the participants articulated an equiprobability answer, we decided to further divide this category of answers into two groups (equiprobability with or without reservation). This choice was made in order to nuance whether the participant questioned equiprobability or else simply assumed it without problematizing it. More explicitly, we classified an answer as *Equiprobable tout court (Group A)* if the text argued without reservation that the outcomes of an 11^{th} coin toss are equiprobable. On the other hand, we classified an answer as *Equiprobable with reservation (Group B)* if the text argued that the outcomes are equiprobable but explicitly expressed some reservation about it. We further classified an answer as *Heads is more likely (Group C)* or *Tails is more likely (Group D)* if the text argued that outcome of an 11^{th} coin toss is more likely to be heads or tails respectively. Finally, we classified an answer as *Mixed (Group E)*, if the text did not conclusively favor any of the above options.

Results

Table 1 summarizes the participants' answers.

Table 1: the participants' answers to Task 1 and Task 2

	Group A Equiprobable tout court	Group B Equiprobable with reservation	Group C Heads more likely	Group D Tails more likely	Group E Mixed	Empty	Total
Yes	3	1	6	0	0	1	11
No	14	13	0	3	0	0	30
In part	15	15	5	0	5	0	40
Not sure	2	1	0	0	0	0	3
Total	34	30	11	3	5	1	84

As to Task 2, we present each group of answers individually together with exemplifying excerpts from the submitted texts which were translated from Italian into English as literally as possible. In presenting the results we pay particular attention to the structure of the arguments expressed and to how these connect to probabilistic concepts and ideas.

Group A

The majority of the participants state that heads and tails are equiprobable without explicit reservation (34 participants). Indeed, most students state that it does not matter what Piero chooses, nor what happened in the first 10 tosses.

> The events of heads and tails are equiprobable. The fact that heads was the outcome […] does not determine a greater probability in the following event.

All the texts categorized in this group are structurally very similar. These students typically state the fact that the events are equiprobable as an unquestioned starting point of their reasoning. This is usually formulated as a statement of the logical-geometric properties of sample-spaces linked to idealized fair coins as described in typical textbooks in probability theory. Other information concerning the previous tosses is then simply dismissed in view of the assumed hypothesis of equiprobability or sometimes not even discussed.

Group B

A consistent group of participants similarly argue that the events of heads and tails are equiprobable. However, these students are careful to explicitly indicate that this depends on the assumption that the coin or the toss is not biased (30 participants). The equiprobability assumption is then given as an explicit (but questionable) hypothesis from which their argument develops. The option of the coin or the game being biased is usually briefly considered as a possibility at the start of the text but dismissed as a result of a deliberate argumentative choice in line with the usual assumptions underlying the practice of problem-solving in probability courses.

> I started from the assumption that the coin was not rigged. The probability […] is the same […] independently from the previous results.

Interestingly, 5 participants explicitly characterize assuming the equiprobability hypothesis as the more "mathematical" way of reasoning.

> Mathematically, the next toss is independent from the previous tosses.

Group C

A smaller group of students argues that an 11th toss is more likely to result in heads (11 participants). This is argued by deeming implausible that a fair toss involving a fair coin could land ten times in a row on heads. Remarkably, 4 participants express this as a contradiction between mathematical thinking and everyday thinking.

> Mathematically the probability for each side is one half. However, since the outcome was 10 times heads, I think that the coin is loaded.

Thus, according to these students, "mathematically" the probability of obtaining head or tails is the same. Nevertheless, if we disregard this, then we can conclude that the coin is loaded.

Group D

Very few students state that an 11th coin toss is more likely to result in tails (3 participants). Interestingly, all these students justify this by giving a mathematical (unsound) argument. For instance, one student argues that if he bets on heads then

> […] Piero has only $1/_{2^{11}}$ probabilities to win.

Group E

Some texts contain considerations on different conflicting aspects of the fictional situation, which they leave unresolved (5 participants). Among these, 3 texts contain reference to unsound mathematical arguments similar to those given by students in Group D.

Discussion

In summary, only a small number of participants shows the positive recency effect (Group C), while a negligible amount manifests negative recency (Group D). A large majority instead inclines towards the equiprobability solution (Group A and B). These results contrast with the findings on university students of Chiesi and Primi (2009) and align better with the results of Fischbein and Schnarch (1997). This happens possibly because the curriculum of studies of our participants is more similar to the curriculum of the students involved in Fischbein and Schnarch's experiment.

As to the participants who manifest the equiprobability effect, their answers show to be connected either directly or indirectly to their education in mathematical probability. The answers of those who do not problematize the equiprobability hypothesis (Group A) may be seen as resulting from applying it as an unquestioned assumption associated with the usual presentation of fictional situations involving idealized games of chance within mathematical textbooks or courses in probability and statistics. As to the students who are keen to problematize the equiprobability hypothesis (Group B and C), they nonetheless state equiprobability as (even explicitly) the more mathematically-proper assumption, i.e., as the assumption which is more appropriate to adopt in a mathematical context. Some of the students even describe an openly-perceived conflict between mathematical and everyday reasoning (a phenomenon also discussed by Rubel, 2007). In particular, positive recency answers and mathematical-probabilistic assumptions were rather presented by the students as conflicting (Group C). On the contrary, the participants who manifested negative recency (Group D) did so when trying to frame the problem in terms of mathematical probability. Given that these were a negligible amount against the total of the participants, we do not attempt to infer more general implications from this particular finding. However, it could be the case that a similar phenomenon is also at work in groups of advanced students in other disciplines who manifest negative recency more consistently (e.g., Chiesi & Primi, 2009).

Thus, we observed an overwhelming tendency by these university students in mathematics to deem the problem's coin to be fair despite indication that this should be the case and contrary to explicit evidence presented in the problem. The equiprobability effect (understood non-normatively) appears to be related to the participants' education in probability, in a way which resembles analogous phenomena reported by socio-cultural qualitative research on word-problems (Verschaffel et al., 2020) and on mathematical proving (Stylianou et al., 2006). Further research testing participants with different mathematical backgrounds using the same procedure would be needed to substantiate this conclusion. Additional research would also be needed to understand if the same students would give substantially different responses when asked the same or an analogous question in a different setting (e.g., in a non-educational setting) or presented by means of a more realistic procedure (e.g., by observing an experiment featuring actual coin tosses). More theoretical elaboration as well as further empirical data would serve in turn to illuminate the relationship between the equiprobability effect and the equiprobability bias.

In conclusion, we have shown how university students in mathematics reason about a problem involving uncertainty. The discussed connection between the equiprobability effect and instruction in probability may be in itself not surprising. Indeed, the equiprobability effect is in general the result of a sound mathematical axiomatization of uncertainty (cf. Gauvrit & Morsanyi, 2014). Some could even argue that the outcome of a formal education in probability should prompt equiprobability answers in all relevant cases. However, the way in which our participants answered show that many of them applied equiprobability as an unquestioned assumption, possibly as a result to concepts and definitions narrowly presented by textbooks and courses in probability and statistics (cf. Batanero, 2020, p. 685). This fact in turn may be problematized as contrary to a full critical understanding of situations involving uncertainty. What if the coin is not fair?

References

Batanero, C. (2020). Probability teaching and learning. In S. Lerman (Ed.) *Encyclopedia of Mathematics Education* (pp. 682-686). Springer. https://doi.org/hh5j.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. http://dx.doi.org/10.1191/1478088706qp063oa.

Chernoff, E. J., & Sriraman, B. (2020). Heuristics and biases. In S. Lerman S. (Ed.) *Encyclopedia of Mathematics Education* (pp. 327- 330). Springer. https://doi.org/hh5h.

Chiesi, F., & Primi, C. (2009). Recency effects in primary-age children and college students using a gaming situation. *International Electronic Journal of Mathematics Education*, *4*(3), 1306–3030. https://doi.org/10.29333/iejme/240.

Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education, 28*, 96–105. http://dx.doi.org/10.2307/749665.

Gauvrit, N., & Morsanyi, K. (2014). The equiprobability bias from a mathematical and psychological perspective. *Advances in Cognitive Psychology*, *10*(4), 119–130. http://dx.doi.org/10.5709/acp-0163-9.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases". *European Review of Social Psychology*, *2*(1), 83–115. https://doi.org/dn8pvc.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychologycal Review, 103*, 592–596. https://doi.org/10.1037/0033-295X.103.3.592.

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. *CBMS Issues in Mathematics Education*, *7*, 234–283. http://dx.doi.org/10.1090/cbmath/007/07.

Morsanyi, K., Primi, C., Chiesi, F., & Handley, S. (2009). The effects and side-effects of statistics education: Psychology students' (mis-)conceptions of probability. *Contemporary Educational Psychology*, *34*, 210–220. https://doi.org/10.1016/j.cedpsych.2009.05.001.

Morsanyi K., & Szucs D. (2014). Intuition in mathematical and probabilistic reasoning. In R. Cohen-Kadosh & A. Dowker (Eds.) *The Oxford Handbook of Numerical Cognition* (pp. 180–200). Oxford University Press.

Paola, D., & Robutti, O. (2001). La dimostrazione alla prova. Itinerari per un insegnamento integrato di algebra, logica, informatica, geometria. *Quaderni del MPI*, *45*, 97–201.

Rubel, L. H. (2007). Middle school and high school students' probabilistic reasoning on coin tasks. *Journal for Research in Mathematics Education, 38*(5), 531–556.

Saenen, L., Heyvaert, M., Van Dooren, W., & Onghena, P. (2015). Inhibitory control in a notorious brain teaser: the Monty Hall dilemma. *ZDM Mathematics Education, 47*, 837–848. http://dx.doi.org/10.1007/s11858-015-0667-6.

Stylianou, D., Chae, N., & Blanton, M. (2006). Students' proof schemes: A closer look at what characterizes students' proof conceptions. In S., Alatorre, J. L., Cortina, M., Sáiz, & A., Méndez (Eds.) *Proceedings of the 28th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education, Volume 2* (pp. 54–60).

Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: a survey. *ZDM Mathematics Education*, *52*, 1–16. https://doi.org/gh45jg.