The influence of the context of conditional probability problems on probabilistic thinking: A case study with teacher candidates

Beyza Olgun, Mine Isiksal-Bostan

The influence of the context of conditional probability problems on probabilistic thinking: A case study with teacher candidates

Beyza Olgun[1] and Mine Isiksal-Bostan[2]

[1]Middle East Technical University, Turkey; bolgun@metu.edu.tr

[2]Middle East Technical University, Turkey; misiksal@metu.edu.tr

This study investigates the influence of the context of conditional probability problems on probabilistic thinking processes of mathematics teacher candidates. Data were collected from four mathematics teacher candidates through semi-structured clinical interviews. Teacher candidates were expected to solve and explain three conditional probability problems, which were selected based on specific contexts. The findings revealed that, when the context is related to a social issue such as health or justice, that can be a hindrance to probabilistic thinking. In addition, while solving questions, teacher candidates focused on their beliefs regarding the social issue rather than using logical and numerical reasoning given in the questions. On the other hand, the findings highlighted the importance of the instruction on conditional probability, as the candidates represented a high-level of probabilistic thinking in the case of consecutive events that were frequently used in curricula.

Keywords: Probabilistic thinking, the context of a conditional probability problem, mathematics teacher candidates

Introduction

As probabilistic thinking involves quantifying uncertainty to arrive at a decision, it has an important role in interpreting uncertain situations in daily life (Amir & Williams, 1999). It helps us better understand whether or not the information gained through our everyday experiences is correct (Metz, 1998). Therefore, probability is recognized as an essential aspect of school mathematics curricula from elementary to undergraduate education (Amir & Williams, 1999). Developing students' skills of probabilistic thinking is one of the main aims of the mathematics curricula in Turkey (MEB, 2018). The development of probabilistic thinking depends on the interactions between "intuition, logical development, and the effects of formal instructions" (Greer, 2001, p. 25). From the educational perspective, the aim is to enhance students' probabilistic thinking from intuitive to more advanced logical and numerical reasoning. However, the instruction of probability concepts in schools may cause discomfort in students' proportional reasoning because probabilistic thinking has a fragile aspect when it intuition is concerned (Fischbein, Nello, & Marino, 1991). This weak aspect may cause students engage in subjective reasoning and to act accordingly or to have some misconceptions. Therefore, while planning their instruction, educators need to consider generating an environment that helps students enhance their conceptual understanding, as well as formalizing their initial intuitive concepts (Greer, 2001).

There are four key constructs regarding probabilistic thinking: sample space, probability of an event, probability comparisons, and conditional probability (Jones, Langrall, Thornton, & Mogill, 1997). Conditional probability is one of the fundamental stochastic ideas that enhance probabilistic thinking throughout the history in Heitele's list (Diaz & Batanero, 2009). It allows individuals to make appropriate changes in their beliefs concerning random events when new information emerges (Diaz & Batanero, 2009). It also helps individuals understand possible risks and make proper decisions in daily life (Watson, 1995). However, compared to the other three key constructs, it requires more careful thinking (Tarr & Lannin, 2005) because of its interest in the outcomes that are elements of a subset of the sample space (Hogg & Tanis, 1993). Therefore, understanding the conditional probability involves the adjustment of the probability of an event when there is a condition regarding the occurrence of another event (Tarr & Lannin, 2005).

Studies showed that students have various misconceptions regarding conditional probability, such as *the transposed conditional fallacy*, *the fallacy of time axis*, and *the base rate fallacy* (e.g., Diaz & Batanero, 2008; Koehler, 1996). The analysis of these misconceptions reveals the influence of the context of probability problems on students' reasoning. In other words, people can approach the problems of conditional probability in various ways when they fail to use appropriate theories (Tversky & Kahneman, 1982). For example, a Bayesian problem that involves two types of information regarding an event – statistics for a population and its specific part of it – can cause a hindrance to student thinking, which leads to the base rate fallacy (Diaz & Batanero, 2009). Moreover, the probability problems that involve medical issues, such as having cancer and taking a related test can be confusing for students as these problems involve the daily experiences of individuals. Students' ways of thinking can be causal based on the scenario or their motive (Tversky & Kahneman, 1982). On the other hand, in some cases where students are accustomed to school experiences, such as consecutive experiments, they can show high-level probabilistic thinking skills (Jones et al., 1997). For instance, the use of representations, such as 2-way tables in the text of a probability problem, could enhance thinking processes of students as well as their achievement (Olgun & Işıksal, 2018).

In this research study, as a continuation of the research study that investigated the role of representations on prospective teachers' thinking processes in conditional probability (Olgun & Işıksal, 2018), the relationship between the context of the probability problems and probabilistic thinking processes is examined. The initial analysis of the previous study pointed out a probable relationship between the context of probability problems and probabilistic thinking processes. Thus, the research question of the present study was as follows: How does the context of probability problems influences prospective mathematics teachers' probabilistic thinking processes in conditional probability based on the framework developed by Tarr and Jones (1997)?

Theoretical Background

Students' understanding of probability, their misconceptions, and their development of probabilistic thinking were the areas of focus in the research literature. To improve the instruction of probability, a coherent frame that reveals students' ways of thinking about probabilistic phenomena is needed (Jones et al., 1997). Therefore, Jones and his colleagues (1997) developed the "Framework for Probabilistic Thinking" to meet this need. The framework consists of four key constructs of probability; *sample space*, *probability of an event*, *probability comparisons*, and *conditional probability*. Tarr and Jones (1997) enhanced this framework for conditional probability. The new framework, which is called the *Framework for Assessing Students' Thinking in Conditional Probability (FASTCP)*, includes *independence* as a new construct. Four levels, namely subjective, transitional, informal quantitative, and numerical are described for each construct of the framework. The following example represents the thinking levels for conditional probability.

There is a cup that contains two red and two blue marbles. Two marbles, one after another, are picked up randomly, without replacement. When students were asked what color marble would be picked given that the first marble is red, based on the FASTCP, the response of "red because it is my favorite color" (Tarr & Jones, 1997, p. 51) would belong to a student in level 1 (subjective). To state it differently, level 1 involves the use of subjective reasoning based on personal traits without considering numerical information. At level 2 (transitional), students begin to use quantitative information, but still, their conditional reasoning is not complete, so students can respond by making such utterances as follows "the probability that blue will be drawn is still the same" (Tarr & Jones, 1997, p. 52). In other words, students misuse numerical information and act with representative heuristics at the transitional level of probabilistic thinking. Moreover, they may still revert to subjective judgments. At level 3 (informal quantitative), students recognize the role of quantitative information for conditional probability as well as monitor the composition of the sample space. Level 3 students can respond by stating that there is one red marble instead of four but still they may also show representative acts or assign incorrect numerical probabilities (Tarr & Jones, 1997). At level 4, as students assign numerical probabilities spontaneously and explain the conditions when the events are related, the response could be that the chance of red is one out of three and the chance of blue is two out of three (Tarr & Jones, 1997). In other words, the numerical level involves the awareness of the composition of the sample space and its essential role in determining conditional probability.

Some studies (e.g., Diaz & Batanero, 2008) showed that students whose probabilistic thinking are at the numerical level might revert subjective reasoning and show biases while performing on probability problems. The importance of the role of intuition in probabilistic thinking and the fragility of probabilistic reasoning in some situations make us inquire what factors influence students' probabilistic thinking. The investigation of students' biases represents the influence of the context of the problems in conditional probability on student thinking. Some studies claim that conditional probability problems, which involve cause and effect relations, confuse students' judgments about the probabilistic phenomena (Biaz & Batanero, 2009). Students rely on causal data that include their perceptions formed via daily life experiences rather than diagnostic data (Diaz & Batanero, 2009). In the case of the Bayesian problems, which involve the issue of base rates, students may ignore the information related to the base rate and tend to rely on the specific information given in the problems (Koehler, 1996). When the context of conditional probability problems involves social contexts, such as justice or health issues, students' beliefs based on personal experiences may get ahead of their logical reasoning. Some studies (e.g., Corter & Zahner, 2007; Watson, 1995) suggest that the use of representations can facilitate students' problem-solving processes and lead them to use logical-numerical reasoning. However, students may still resort to erroneous judgments for problems based on certain contexts rather than using representations as a model (Olgun & Işıksal, 2018). Based on the literature it could be deduced that there are limited studies that have focused on the context of probability problems in the case of conditional probability while modelling and representations have been investigated by many researchers (Chaput, Girard & Henry, 2011). These issues necessitate a study investigating how different contexts of probability problems influence probabilistic thinking processes. Such a study can reveal some information about the instruction of conditional probability as well as students' thinking processes. Therefore, the current study will investigate the influence of different contexts in conditional probability problems on probabilistic thinking processes of teacher candidates based on the FASTCP.

Methodology

Gaining a deep understanding of participants' thinking processes was the main objective of the present study, in which semi-structured clinical interviews were held with the participants, who were selected by means of the convenient sampling method. The participants were students, who were in their 4th semester of teacher education program and who had taken "MATH 112 Discrete Mathematics" and "STAT 201-202 Introduction to Probability and Statistics I-II" courses. These students were informed about the research study. The interviews were conducted with the volunteer students. Four (two male and two female second-year students between the age of 20-22 years) students enrolled in the mathematics teacher education program of one of the largest public universities in Turkey constituted the participants of the study.

In the interviews, the participants were asked to perform on three specific conditional probability problems (see Figure 1), which were selected from the Conditional Probability Reasoning Test (CPR) developed by Diaz and Batanero (2009). Each problem was determined according to its specific context. Problem 1 was related to a consecutive experiment involving computing conditional probability and dependence of the events. Problems 2 and 3 included the social context of justice and health issues, respectively. In problem 2, which is a Bayesian problem, two types of information were given: the frequency of the base rate and specific information about the case. Problem 3 was related to a causal issue including the application of a medical test and whether or not there was an illness.

Problem 1: (Falk 1986). Two black and two white marbles are put in an urn. We pick a marble from the urn. Then, without putting it back into the urn, we pick a second marble at random.

If the second marble is white, what is the probability that the first marble is white? $P(W_1 | W_2)$

i. 1/3
ii. Cannot be computed
iii. 1/6
iv. 1/2

Problem 2: (Tversky & Kahneman 1982a). A witness sees a crime involving a taxi in a city. The witness says that the taxi is blue. It is known from previous research that witnesses are correct 80% of the time when making such statements. The police also know that 15% of the taxis in the city are blue, the other 85% being green. What is the probability that a blue taxi was involved in the crime?

a. $\frac{80}{100}$

b. $\frac{15}{100}$

c. $\frac{15}{100} \times \frac{80}{100}$

d. $\frac{15 \times 80}{85 \times 20 + 15 \times 80}$

Problem 3: (Pollatsek, et al. 1987). A cancer test is administered to all the residents in a large city. A positive result is indicative of cancer and a negative result of no cancer. Which of the following results is more likely or are they all equally likely?

a. A person has in fact cancer supposed that he got a positive result.
b. To have a positive test supposed that the person has cancer.
c. The two events are equally likely.

Figure 1: The problems, adapted from "University Students' Knowledge and Biases in Conditional Probability Reasoning", by Diaz and Batanero (2009, pp. 157–158)

The participants were asked to submit their solutions both in written form and verbally, and the first author asked related questions to understand their in-depth thinking processes. Each interview lasted 90 to 120 minutes and was videotaped and transcribed. The demographic information and the consent of the participants were obtained in advance. In the data analysis process, students' probabilistic thinking processes were explored based on the FASTCP for each problem. Two researchers performed open coding. After the codes were determined individually, the inter-rater agreement was calculated to be 90%. Different ideas were discussed until a consensus was reached. General codes were identified based on their match with the respective thinking levels from the FASTCP. For instance, the general codes of "assigning numerical probabilities spontaneously and accurately" and "using and explaining logical-numerical reasoning" were matched with level 4; the general code of "reverting to subjective reasoning even though recognizing numerical information" was matched with level 2. Expert opinion was also obtained on the appropriateness of matching the general codes with the thinking levels.

Findings

The analysis of the data showed that the context of the problem had an apparent influence on probability thinking processes of the mathematics teacher candidates. It was observed that when the conditional problem includes a social context, which was, in this case, a health issue or justice for a crime, the participants focused on their beliefs based on their daily life experiences and answered the question accordingly. In other words, most of the time, they resorted to subjective reasoning rather than to logical and numerical reasoning. On the other hand, they presented high-level probabilistic thinking in consecutive experiments.

In problem 1, all participants represented the skills of the numerical level of probabilistic thinking. All participants easily recognized the condition for the consecutive experiment. They were aware of the composition of the sample space, which plays an essential part in determining the conditional probability, and applied numerical probabilities with relative ease when the other situations are considered. They could state the necessary conditions whether or not two events were related. They used numerical reasoning while comparing the probabilities of the events for each trail in non-replacement situations. In the following example, it can be seen that the participant executed the required conditions for conditional probability and assigned precise values for the desired events.

Student C: The probability that the first marble is white if the second marble is white... First I will find the probability that the second marble is white. These are WW (white and white) or BW (black and white). For the situation of WW, the probability is $\frac{1}{2} \times \frac{1}{3} = \frac{1}{6}$. For the case of BW, the probability is $\frac{1}{2} \times \frac{2}{3} = \frac{1}{3}$. Then $\frac{1}{6} + \frac{1}{3} = \frac{1}{2}$ is the probability that the second marble is white. Here, the probability that the first marble is also white is $\frac{1}{6}$. Then, as like before I will calculate the proportion of $\frac{1}{6}/\frac{1}{2}$. It is $\frac{1}{3}$.

In problem 2, only one participant displayed the skills of numerical level thinking. The other participants displayed the transitional level of thinking. They presented acts of representativeness as a confounding effect when they made decisions and ignored Bayesian reasoning in their solutions.

They confused the situation of “the probability of the involvement of a blue taxi in the crime" with "the chance of finding the taxi that was involved in the crime if it is blue”. Therefore, they were confused because the possibility of the taxi involved in the crime being blue should be higher as there was a witness when their numerical calculations said the opposite. Another participant also had the misconception of the base rate fallacy. He ignored the specific information regarding the witness and only focused on the frequency of blue taxis as can be seen in the following example.

Student A: Is it not an easier thing for us to have 15% blue? Obviously, the thing that blue taxis are 15% of the taxis in the city increase the possibility to find the taxi that was involved in the crime because the less the rate, the more likely it is to find. ... The possibility of telling the truth is 80%. Then the possibility of being blue is 15% but as I said it is a 15% chance of blue from all taxis. For example, here it multiplies 80/100 with 15/100 but because I know that 15% of the taxis are blue if I search within this 15%, the possibility of finding will be high. ... I will say that the possibility of finding the taxi that was involved in the crime is one in fifteen...

In problem 3, all participants acted in the transitional level of thinking. They misused numbers in determining probabilities. For example, when the sample space contained two outcomes as positive and negative results, they assumed that these outcomes are equally likely. Students did not use any representation for the information in the problem to facilitate their solutions. Their judgment about the conditional probability reverted to subjective reasoning as can be seen in the following episode:

Student A: In the case of a positive result when the person has cancer, in the case of having cancer when the result of the test was positive… I'm not a person who trusts in tests that related to health, so I think that the probability of having a positive result in the state of being ill will be lower.

Interviewer: It was a subjective judgment; do not you think?

Student A: Yes, but the question is also very subjective

Interviewer: Can we make a numerical comparison?

Student A: If I have this disease, the possibility of my being patient according to the test result %50 likely. I assume I have cancer then there are two possibilities that this device will show me as plus or negative. If not, it is the same for the other also… So, as I said, I still think that the possibility of having a positive test result supposed that the person has cancer will be higher than the probability of the person has in fact cancer supposed that the person has cancer.

Discussion

The findings of the present study revealed the fragile aspect of probabilistic thinking. Teacher candidates' actions varied depending on the context of conditional probability problems. Their intuition, which is one of the main components of probabilistic thinking (Greer, 2001), came into prominence when the context was related to a social issue. On the other hand, they showed high-level probabilistic thinking in consecutive experiments. While they represented all the outcomes via listing or creating a tree diagram for consecutive experiments, they did not use any representation for the other two problems. Therefore, they could not explain their reasoning within logical boundaries based on data representation for the problems associated with social issues. For instance, they could not express the need for the accuracy of the cancer test to solve problem 3. The main reason for this difference might be the standing of the problems related to consecutive experiments in school curricula and the instruction of probability. Consecutive experiments have been frequently used in textbooks as well as in regular instructions. Therefore, teacher candidates may have the appropriate approach in such scopes. Thus, they may directly focus on logical reasoning and numerical probabilities without resorting to their intuition. On the other hand, in a social context, they resort to subjective reasoning or ignore the importance of the condition.

The frequentist approach that recommends a sequence, which starting with the introduction of data followed by probability through datasets (Chaput et al., 2011), has been implemented in the Turkish school curricula since 2009. Therefore, our participants' initial acquaintance with probability was by means of the old approach, which starts with probability and introduces data environment through chance situations. Thus, the introduction of the conditional probability might have been difficult in adopting probability teaching based on data in high school years. Advanced mathematics courses in undergraduate studies have also been criticized for their instructional methods and the degree to which the required conceptual change regarding the related content is created (CBMS, 2011). These issues highlight the essential role of the instruction of probability. Teachers should enhance students' probabilistic thinking regardless of the context. Therefore, teachers should include both consecutive experiments and probability problems with different contexts in their instruction of conditional probability, rather than continue using similar contexts. Students should adopt logical and numerical reasoning in different contexts, which are used for consecutive events. They should focus on the constructs of probability problems, which help to form conditional probabilistic reasoning, rather than on contextual issues. These suggestions point out adopting a frequentist and problem-solving approach in the context of real data for the instruction of probability (Shaughnessy, 2003).

The current study is limited to the contexts of consecutive experiments and social issues of justice and health. Further studies should be conducted for the comparison of different contexts of probability problems to gain an in-depth understanding of the influence of the context on probabilistic thinking. Moreover, the investigation of how experts use logical and numerical reasoning in different contexts might provide significant insight into the instruction of conditional probability.

References

Amir, G. S., & Williams, J. S. (1999). Cultural influences on children's probabilistic thinking. *The Journal of Mathematical Behavior, 18*(1), 85–107.

Chaput, B., Girard, J. C., & Henry, M. (2011). Frequentist approach: Modelling and simulation in statistics and probability teaching. In C. Batanero, G. Burrill, and C. Reading (Eds.), *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 85–95). Dordrecht: Springer.

Conference Board of the Mathematical Sciences [CBMS] (2001). *The mathematical education of teachers* (Vol. 11). Providence, RI: American Mathematical Society.

Corter, J. E., & Zahner, D. C. (2007). Use of external visual representations in probability problem solving. *Statistics Education Research Journal*, *6*(1), 22–50.

Diaz, C., & Batanero, C. (2008). *Students' biases in conditional probability reasoning*. Paper presented at ICME 11 (International Congress on Mathematical Education): TSG 13, Monterrey, Mexico.

Diaz, C., & Batanero, C. (2009). University students' knowledge and biases in conditional probability reasoning. *International Electronic Journal of Mathematics Education*, *4*(3), 131–162.

Fischbein, E., Nello, M. S., & Marino, M. S. (1991). Factors affecting probabilistic judgments in children and adolescents. *Educational Studies in Mathematics, 22*(6), 523–549.

Greer, B. (2001). Understanding probabilistic thinking: The legacy of Efraim Fischbein. *Educational Studies in Mathematics*, *45*(1), 15–33.

Hogg, R. V., & Tanis, E. A. (1993). *Probability and statistical inference* (4th ed.). New York, NY: Macmillan.

Jones, G. A., Langrall, C. W., Thornton, C. A., & Mogill, A. T. (1997). A framework for assessing and nurturing young children's thinking in probability. *Educational Studies in Mathematics, 32*(2), 101–125.

Koehler, J. J. (1996) The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavior and Brain Sciences, 19*(1), 1–54.

Metz, K. E. (1998). Emergent ideas of chance and probability in primary-grade children. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (pp. 149–174). New York, NY: Routledge.

Milli Eğitim Bakanlığı [MEB] (2018). *Orta öğretim matematik dersi (9, 10, 11 ve 12. sınıflar) öğretim programı* [Secondary education (Grades 9-12) mathematics curriculum]. Turkey: MEB. Retrieved from http://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=343.

Olgun, B. & Işıksal, M. (2018, October). *The role of two-way tables in conditional probability problems: The case of prospective teachers*. Paper presented at The European Conference on Educational Research: 24. Mathematics Education Research, Bolzano, Italy.

Shaughnessy, J. M. (2003). Research on students' understanding of probability. In J. Kilpatrick, W.G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics*, (pp. 216–226). Reston, VA: NCTM.

Tarr, J. E. & Jones, G. A. (1997). A framework for assessing middle school students' thinking in conditional probability and independence. *Mathematics Education Research Journal, 9*(1), 39–59.

Tarr, J. E. & Lannin, J. K. (2005). How can teachers build notions of conditional probability and independence? In G. A. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning*, (pp. 216–238). New York, NY: Springer.

Tversky, A. & Kahneman, D. (1982). Judgements of and by representativeness. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases,* (pp. 84–98). New York, NY: Cambridge University Press.

Watson, J. M. (1995). Conditional probability: Its place in the mathematics curriculum. *The Mathematics Teacher*, *88*(1), 12–17.