INTEGRATING THEORETICAL AND EMPIRICAL CONSIDERATIONS - YOUNG STUDENTS' UNDERSTANDING OF THE EMPIRICAL LAW OF LARGE NUMBERS

Susanne Schnell
University of Paderborn, Germany
susanne.schnell@math.upb.de

In German middle schools, probabilities are first approached either through theoretical considerations (e.g. by comparing the number of ways an event can occur with the total number of possible outcomes) or empirical investigations (e.g. by conducting series of experiments). The empirical law of large numbers serves as a means to achieve an integrated view of both approaches. Understanding this law is challenging for young learners as it draws on a complex network of statistical concepts. This paper investigates students' processes of experiencing the empirical law of large numbers. Interpretatively analyzed snapshots from a design research study with students (age 11-13) illustrate how they gradually integrate empirical and theoretical considerations when dealing with a teaching-learning arrangement built on a computer simulation.

THEORETICAL AND EXPERIMENTAL APPROACHES AND THEIR INTEGRATION

In school, probabilities can be determined by different approaches: For instance, a 20-sided die with differently colored sides is rolled once and students are asked to determine the probability of the event "red side up". In the so-called classical approach, students will determine the theoretical probability by considering the sample space: counting the number of sides with the color red (e.g. seven red sides) and the overall number of sides (20 sides) and determine the ratio of both values as P(red side up) = 7/20 = .35. However, this specific approach is only applicable under the assumptions that the die is fair and each of its sides has the same surface area and thus the same likelihood to be facing up (1/20). For a secretly manipulated die which looks normal, the same considerations could be made but lead to wrong results – which would possibly only become apparent if one or more series of experiments were conducted. Throwing the die is not necessary, though, as long there is no doubt about its fairness.

Another typical approach is the empirical or experimental approach: To determine the probability of an event, an experiment is conducted repeatedly and the results are noted. By determining the relative frequency of a long series of repetitions (or the average frequency of many long series), the probability is estimated. The value of this approach in school is especially apparent when theoretical considerations are impossible, for instance for very irregular dice.

The experimental and theoretical probabilities are closely related: "The relationship between the two concepts results from the fact that for a given event, experimental probability will more closely approximate theoretical probability as the number of trials increases" (Jones et al, 1999, p. 148). This so-called empirical law of large numbers makes a statement only concerning long series of repetitions of experiments. As Konold (1989) showed, students often tend to focus on interpreting probabilities in regards to single outcomes, which is rather difficult and possibly not very valuable: if the colored icosahedron with the color distribution described in figure 1 is rolled once, the red side is the most likely (as the probability of each other color is smaller), but still rather unlikely (as it has only a probability of .35). A more specific statement can only be made when predicting the relative frequency in a large series of events. Thus, the empirical law of large numbers is crucial for a fundamental understanding of probabilities: "The empirical law of large numbers explains why one can adopt probabilistic conceptions in a successful way although random cannot be calculated for single outcomes. It explains the sense and the preconditions, but also the limits of probabilistic considerations" (Prediger, 2008, p. 16). Understanding this relationship between theoretical and empirical considerations is challenging for young learners (Jones et al., 1999). Especially relating the observability of patterns to different sample sizes poses a conflict (Ireland & Watson, 2009).

CHALLENGES FOR AN INTEGRATED APPROACH

Therefore, questions arise how to introduce probabilities in school so that students will eventually be able to grasp both concepts and get an integrated understanding of them. According

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

to Konold et al. (2011) commonly used approaches often have "the goal that students come to expect that the relative frequency of actual trials of some chance phenomenon will converge to the theoretical probability as the number of trials grows large" (p. 70). In this case, the data analysis in order to determine the relative frequency is only looking for more or less stable patterns which fit the expectations raised by the theoretical considerations. The variability of the data is an unwanted and possibly disregarded noise (cf. Riemer, 1991, p. 18). One source of this problem is the over-confidence in the theoretical approach. Probabilities from each approach serve only as models for the true, unknown and unknowable probability of the real situation. It is wrong to identify each one of them as the true probability itself (cf. Konold et al., 2011). Probabilities are inherently hypotheses, which have to be replaced in case of the availability of a more appropriate hypothesis (cf. Riemer 1991, p. 19).

If this hypothetical character is lost, it can be the root of conceptual problems: "When students think from the start that they know the true probability (i.e., that it is the theoretical probability), then the idea of uncertainty and confidence in one's inference—indeed, the very idea of an inference—is lost" (Konold et al., p. 83). To address the variability which causes the uncertainty it is better to compare different samples with each other rather than a successively growing sample (cf. Freudenthal, 1972). "From the multiple repetitions, the students get a better sense of sample-to-sample variability (…) we cannot speak or think about data having 'signal' without simultaneously thinking or speaking about the noise and vice versa. The two ideas are co-constructed" (Konold & Kazak, 2008, p. 30). This interplay of patterns (signals) and variability (noise) is at the heart of random phenomena (Moore, 1990, p. 135).

Overall, to acquire a substantial and integrated understanding of probabilities, students have to coordinate different concepts such as theoretical and experimental perspectives, variability and patterns in data and the influence of the sample size. For this, computer simulations are a crucial tool in order to facilitate the analysis of very large samples (cf. Konold et al., 2011). While Konold et al. (2011) suggest beginning school instruction with experiments for which theoretical considerations are not possible, another possibility is to begin with a theoretical perspective and then focus on its power and limitations when it comes to predicting outcomes in data. Using this latter approach, this paper will investigate students' learning pathways in regards to the following question: *How do students informally coordinate theoretical and empirical aspects when they are aware of the theoretical sample space?*

DESIGN AND SETTING

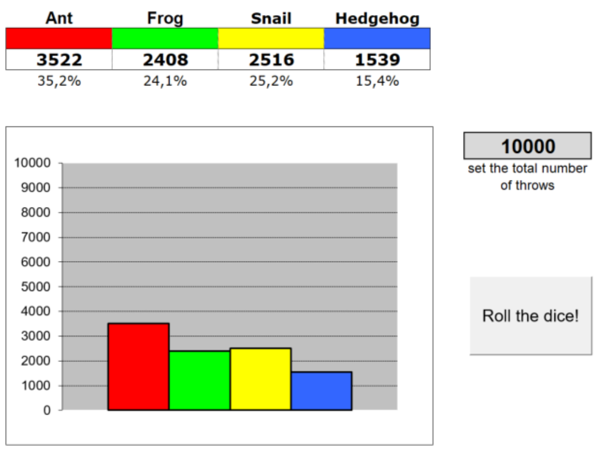
Teaching-learning arrangement 'Betting King'

The interview data presented in this paper come from a design research project which is constructed around a teaching-learning arrangement called 'Betting King'. A condensed version of this is published as chapter of a German textbook for mathematics classrooms in grade 7 (age 12-13) (Leuders et al., 2015). In the teaching-learning arrangement, students are asked to investigate a game, in which a 20-sided die with four different colors is rolled in order to move accordingly colored animals in a race. The race ends after a previously determined number of rolls which can differ between 1 and 40 for the board game and between 1 and 10.000 in the later used Excel simulation. In the initial game phase, students are asked to make a bet on the animal they think will have moved the farthest after the determined number of rolls. Betting on the correct animal gives the player a point and the person with most points becomes the 'betting king'. After playing the game first completely freely and then supported by record sheets determining the lengths of each game (cf. figure 2), students enter the investigation phase. Here, they are asked to find a *good* betting strategy and a way to make a bet *as secure as possible*. Speaking in terms of a theoretical and experimental perspective, the first task leads to analyzing the theoretical probability while the second intends to focus on the sample size of experiments.

An analysis of the die reveals that the color distribution is in favor of the red ant with a probability of .35 in each throw (cf. figure 1), which makes it theoretically the best bet. To make a bet as secure as possible, the length of the game has to be taken into account as well: For longer games, the red ant is more likely to win every time while in short games other animals can be perceived to win occasionally. To investigate these questions, students use the Excel simulation, which generates the result of a game of any length immediately (i.e. the progression of the game is not shown but the result is portrayed as absolute (and later in the design experiments also relative) frequencies for each animal as well as bar charts; cf. fig. 1)). In these investigation phases, record sheets are used to structure the data collection and analysis and facilitate the focus on differing sample sizes. In preliminary studies (Prediger & Rolka, 2009), the focus on the differing lengths of games ("short" vs. "long" games, which is deliberately a soft description and its definition can vary between students) proved to be especially challenging for students, but crucial to gain insights into more or less stable patterns (in the long run) and a high variability (in the short run) (similar Pratt et al., 2008, p. 127). While most students figured out the color distribution and thus the sample space for a theoretical consideration of chances relatively early in the design experiment, others assumed the die to be fair or disregarded the distribution at all, basing all bets only on the empirical results. In this case, students were asked explicitly towards the end of the first design experiment session to look more closely at the die.

Animal	# of sides	Probability
Red ant	7	0.35
Green frog	5	0.25
Yellow snail	5	0.25
Blue Hedgehog	3	0.15
Total	**20**	**1**

Figure 1. Overview of animals, color distribution of sides on the die and theoretical probability



Total number of throws	I bet on...	Winning animal	points
1	*Red ant*	*Red ant*	*1*
2	*Red ant*	*Red ant, green frog*	*1*
5	*Red ant*	*Red ant, green frog*	*1*
10	*Red ant*	*Yellow Snail*	*0*
20	*Red ant*	*Red ant*	*1*
		Total number of points:	*4*

Total number of throws	I bet on...	Winning animal	points
1	*Green frog*	*Green frog*	*1*
2	*Green frog*	*Red ant, yellow snail*	*0*
5	*Green frog*	*Red ant, yellow snail*	*0*
10	*Green frog*	*Red ant, yellow snail, blue hedgehog*	*0*
20	*Red ant*	*Red ant, green frog*	*1*
		Total number of points:	*2*

Figure 2. Left: display of the Excel simulation; right: first record sheet of games played on the board, italic writing was filled in by the students John and Victor (translated by the author); later record sheets systematically increase up to the total number of 10,000 throws

In the third phase, students explore more closely the frequency of each animal in relation to the (high) total number of rolls and the color distribution of the die. The intention is to further specify the observable patterns (e.g. 'when I simulate a game of 2,000 throws, the red ant is very often reaching between 650 and 750 throws or approximately 35 %'). This phase will not be addressed in this paper due to the page limitations (for more information cf. Prediger & Schnell, 2014).

The main strategy for investigation supported by the structure of the Excel simulation is the comparison of many game results with the same total number of throws to identify patterns. In a so-called static-comparative view (Schnell, 2014), observations made in a series of games of a specific length (e.g. 2,000) are then compared with observations in a series of games with a different length (e.g. 20). For a final investigation phase, the Excel simulation also offers a dynamic view with a gradually growing number of throws, but most children did not investigate this in detail.

Data collection and analysis

The teaching-learning arrangement 'Betting King' was used in design experiments (cf. Cobb et al., 2003) with nine pairs of students (age 11-13) in a German comprehensive middle school who had not been introduced to probabilities in mathematics lessons. Each pair worked on the teaching-learning-arrangement in four to six consecutive lessons of normally 90 minutes each. The collected data consists of videotapes of all sessions (40 in total), screen captured videos of the simulation activities and all written products such as record sheets.

For the in-depth analysis, all sessions of two pairs of students as well as selected sessions or scenes from other pairs were transcribed. The analysis was conducted in groups of researchers and followed the interpretative paradigm, aiming at reconstructing students' processes of conceptual development (cf. Schnell, 2014 for more details). Initially, the data was scrutinized for relevant scenes in which students discussed new ideas or offered explanations. Next, these scenes were sequentially interpreted in order to reconstruct the students' inner logic (e.g. 'why are they saying what in regard to which situation?'). Lastly, the students' statements were interpreted in regard to theoretical and empirical considerations.

RESULTS

The following excerpts are taken from the design experiment session of two pairs who identified the color distribution of the 20-sided die relatively early in the investigation. Presented are core scenes in which the students' coordination of theoretical and experimental consideration becomes apparent.

John and Victor – gaining confidence in a theoretically founded strategy by experiments

Very early into the design experiment (after 3 games), John assumes that there is more red on the die. John verifies his conjecture by counting the sides on the die, but arrives at "seven times red, all others six each". Consequently, the students bet only on the red ant in the following games, which can be interpreted as theoretically founded strategy. However, while the red ant wins four out of five games (cf. figure 2, upper block), the boys start to use the record sheet for a systematic investigation by choosing a different animal to bet on per block of five games (length 1-20). Instead of either counting the number of times the red ant or each of the other animals won overall in these experiments, John verbalizes a strategy which relies solely on different empirical considerations:

John: *But maybe I now know the strategy: one block, bet on each animal and when you know which of them won most often, then you bet on it in the 20s game.*

This can be understood as a strategy of collecting data and inferring the 'best' animal out of four games. The way John puts it, his initial considerations of the red ant's superiority are not taken into account. At this point, the length of the games is completely disregarded by the boys. However, as they have only played short games of up to 20 throws, the winning animals vary a lot which might possibly lower their confidence in the red ant's favor.

When introduced to the computer simulation, the boys immediately start to explore games with a length of 10,000 throws:

Victor: *The [blue] hedgehog['s bar in the bar chart] always goes up and down a bit.*
John: *[Red] ant wins always.*
Victor: *No wonder! (…)*
John: *So, [green] frog and [yellow] snail are always the same, [red] ant wins always and [blue] hedgehog loses always.*

Victor first observes the variability in the changing bar charts from game to game. John on the other hand focusses on patterns in regard to the order of animals and uses the term 'always' to indicate their stability. However, when introduced to a record sheet with four games of length 1, 10, 100, 1000 and (added by the students) 10,000 each, the boys do not apply this observation in their betting. Again, they revert to their investigative strategy of betting systematically on a different animal in every game. In the second record sheet of the same kind, they switch to betting always on the red ant, which they conclusively evaluate as "super successful". Here, it seems the students build confidence in the red ant from an experimental perspective.

At the end of phase 2 of the design experiments, the students determine which animal for each number of throws and a final strategy is noted: “You can bet most securely at 10,000, 1,000 and 100, because the higher the number, the bigger the chance to win”. Asked to clarify which animal they would choose for a most secure bet, the immediately state:

Victor: [Red] ant, of course. It wins always.
John: No, only for high numbers. Because it has more red sides.
Victor: Yeah, I meant that. More red sides and wins more.

Coming from a theoretical perspective, the students do not use the theoretical superiority of the red ant in their betting but rather use the simulation for explorations. Their first strategies seem to be solely empirically founded. When switching to long games, their confidence in the red ant is increased and they finally integrate the sample size and the color distribution.

Elisa and Jacob: Interpreting empirical results through the lens of theoretical considerations

Elisa and Jacob immediately begin by betting on the red ant in the first and all subsequent free games, which also wins every time (length between 20 and 40). While the researcher explains the record sheet, the students begin to count the sides on the die.

Researcher: What did you count?
Elisa: Yellow. (…) Maybe how fast the animals are? (quietly) As much as they have [sides on the die]?
Jacob: Yes, [red] ant is quite fast.
Elisa: Yes, maybe because there are many reds.

The theoretical consideration according to the color distribution is applied to the ‘speed’ of the animals and thus raises an expectation for the results of games. Consequently, Elisa and Jacob bet in each game of the first type of record sheets (cf. fig. 2) on the red ant, winning between 2 and 4 points per block. As they seem very confident in their strategy, the researcher points out a game in which the students didn’t win a point.

Researcher: But the ant lost that time. Do you want to choose another animal?
Jacob: No, of course not. It lost because it’s just a game of chance.(…) It’s still the fastest!

Jacob explains the outlier with “chance” while sticking to his evaluation of the red ants superiority. Overall, even the short games seem to enhance the students’ confidence in the ant which is rooted in the theoretical considerations.

When using the computer simulation, the ant loses the first four games with length 1. Thus, Jacob suggests choosing another animal, but when the red ant wins again (at game length 10), the students change back and stay with the red ant. Jacob then begins to explore games with length 1,000 “to see, if the [red] ant wins every time”. Both students are surprised when no other animal than red wins. Evaluating their strategy of betting on red ant, they determine

Jacob: We are better [with the bet on red ant] at 100s and 1000s than anywhere else.
Elisa: Yes, better at 100s. (…) We are better at the end.
Jacob: (...) [most secure is] 1,000 because (...) – well, if I consider my *experiences*- our *experiences, then the red ant won always*
Jacob writes “You can bet the most securely at 1,000, because ~~you have more throws~~ our animal won every time”
Elisa: We knew it. More red sides mean the ant is fast. Very fast.

Even though the students considered the red animal to be the best bet the whole time, they quickly assume that the bet is more secure for longer games according to their empirical experiences. Interestingly, Jacob emphasizes the subjectivity of this experience (“*my* experiences”), which could be an indicator of an insecurity if the sole win of the red ant for games of length 1,000 would occur for other players or if a longer exploration phase would show other animals winning.

Overall, Elisa and Jacob begin very confident in the red ant due to a very early determination of the color distribution. Data from short and long games is evaluated mostly on the backdrop of the expectation of the ant being the fastest. Thus, games in which it loses are regarded as “chance” outliers which do not diminish the superiority of red. However, the surprised reaction to red ant being the sole winner in a series of games of length 1,000 indicates that Elisa and Jacob still expect variability in the winning animals as they perceived it in the short games. They quickly consider longer games as more secure to bet based on their empirical experience which Elisa in her final statement relates to the theoretical chance.

DISCUSSION

The presented snapshots from the design experiments illustrate how children who discover the color distribution and interpret it correctly in terms of (theoretical) chances deal with subsequently collected data. The first pair seems less confident and uses the data to back up their initial considerations, which is complicated by the variability of results. The second pair interprets all data as either confirmation of the theoretical chances or as 'chance' outliers. Thus, not only the confidence in the expected pattern is different, but also the way they deal with variability in data. In line with Konold & Kazak (2008) students not only construct insights into theoretical and empirical probability, but also into the interplay of patterns and variability. Regarding the hypothetical status of probabilities (Riemer, 1991), both pairs account for an insecurity of their red-ant-winning-strategy but the first pair is more likely to investigate and apply different strategies based on empirical data. Even though the sample space is known to the students, the interpretation of data and thus the coordination of theoretical and empirical considerations are delicate. However, the presented study shows promise that students are able to draw these connections.

REFERENCES

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher, 23*(1), 9-13.

Freudenthal, H. (1972). The Empirical Law of Large Numbers, or the Stability of Frequencies. *Educational Studies in Mathematics, 4,* 484-490.

Ireland, S., & Watson, J. (2009). Building a connection between experimental and theoretical aspects of probability. *International Electronic Journal of Mathematics Education*, *4*(3), 339-370.

Jones, G., Thornton, C., Langrall, C., & Tarr, J. (1999). Understanding Students' Probabilistic Reasoning. In L.V. Stiff and F.R. Curcio (Eds.), *Developing Mathematical Reasoning in Grades K-12: 1999 Yearbook* (pp. 146–155). Reston, VA: National Council of Teachers of Mathematics.

Konold, C. (1989). Informal Conceptions of Probability. *Cognition and Instruction*, *6*, 59-98.

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education, 2*(1), Article 1.

Konold, C., Madden, S., Pollatsek, A., Pfannkuch, M., Wild, C., Ziedins, I., Finzer, W., Horton H. J., & Kazak, S. (2011). Conceptual Challenges in Coordinating Theoretical and Data-centered Estimates of Probability. *Mathematical Thinking and Learning, (13)*1-2, 68-86.

Leuders, T., Prediger, S., Barzel, B., & Hußmann, S. (Eds.). *Mathewerkstatt 7*. Berlin: Cornelsen.

Moore, D. S. (1990). Uncertainty. In L. S. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95-137). Washington, DC: National Academy Press.

Pratt, D., Johnston-Wilder, P., Ainley, J., & Mason, J. (2008). Local and Global Thinking in Statistical Inference. *Statistics Education Research Journal*, *7*(2),107-129.

Prediger, S. (2008). Do you want me to do it with probability or with my normal thinking? Horizontal and vertical views on the formation of stochastic conceptions. *International Electronic Journal of Mathematics Education*, *3*(3), 126-154.

Prediger, S., & Rolka, K. (2009). Using betting games for initiating conceptual change - A case study in probability classrooms. *Asian Journal of Educational Research and Synergy*, *1*(1), 57-67.

Prediger, S., & Schnell, S. (2014). Investigating the Dynamics of Stochastic Learning Processes: A Didactical Research Perspective, Its Methodological and Theoretical Framework. Illustrated for the Case of the Short Term–Long Term Distinction. In E. J. Chernoff & B. Sriraman (Eds.). *Probabilistic Thinking: Presenting Plural Perspectives (pp. 533-588)*. Dordrecht: Springer.

Riemer, W. (1991). *Stochastische Probleme aus elementarer Sicht*. Mannheim/Wien/Zuerich: BI-Wissenschaft Verlag.

Schnell, S. (2014). *Muster und Variabilität erkunden – Konstruktionsprozesse kontextspezifischer Vorstellungen zum Phänomen Zufall*. Wiesbaden: Springer Spektrum.