THE EFFECT OF VISUALIZING STATISTICAL INFORMATION IN BAYESIAN REASONING PROBLEMS

Markus Vogel[1] and Katharina Böcherer-Linder[2]
[1]Institute for Mathematics and Informatics, University of Education Heidelberg, Germany
[2]Institute for Mathematics Education, University of Education Freiburg, Germany
vogel@ph-heidelberg.de

In this paper, we compare the effectiveness of two visualizations, i.e. the tree diagram and the unit square, in increasing people's performance in Bayesian reasoning tasks. We summarize the results of three experiments with randomized samples of about nearly 300 university students. In the first experiment the students only had to compute ratios of different set-subset relations included in Bayesian situations. On step further, in the second experiment the students had to solve tasks by applying Bayes' rule. In the third experiment we investigated the visualizations concerning the students' ability to assess the impact of changing base rates in Bayesian situations. In all experiments there was empirical evidence for the unit square being more supportive. On this base we conducted a replication study with about 140 students in six different classes of lower secondary level to research on potential effects of mathematical education. The survey and the data-analysis of the fourth study are taking place currently. The results will be presented at the conference.

INTRODUCTION

The difficulties people usually have when coping adequately with Bayesian reasoning situations are widely known. Cosmides and Tooby (1996) speak of "clashes between intuition and probability" which are in line with the fact that even experts are mainly failing at Bayesian based judgements in their field of expertise by underlying "cognitive illusions" (Tversky & Kahneman, 1974). Different researchers, for example, Eddy (1982) or Hoffrage and Gigerenzer (1998) report on studies in which only maximally 10% of the participating physicians were able to interpret results of a medical diagnosis like the following in the right way:

> "10% of women at age forty who participate in a study have a particular disease. 60% of women with the disease will have a positive reaction to a test. 20% of women without the disease will also test positive. " (Johnson & Tubau, 2015, p. 3)

Figure 1. A Bayesian situation referring a medical diagnosis test.

Because misinterpretations in such situations of decision making could have serious consequences huge efforts in research have been made to identify difficulties in Bayesian reasoning situations and to look for strategies how to overcome them. In the meanwhile, psychological and educational research found out that the question of the Bayesian problems' representation is crucial with regard to the kind of numerical information on the one hand, and to the kind of visualization on the other (McDowell & Jacobs, 2017).

Representing numerical information via natural frequencies instead of probabilities facilitates Bayesian reasoning in situations like a medical diagnosis (e.g. Hoffrage & Gigerenzer, 1998; Binder, Krauss, & Bruckmaier, 2015). In terms of this approach, the problem is given by:

> "10 out of 100 women at age forty who participate in a study have a particular disease. 6 out of 10 women with the disease will have a positive reaction to a test. 18 out of 90 women without the disease will also test positive."

Figure 2. The medical diagnosis test situation represented by natural frequencies.

The correct solution, which is the so-called positive predictive value of the test, can be calculated by Bayes' rule either by using probabilities (cf. Fig. 1) or natural frequencies (cf. Fig. 2):

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

$$P(disease|positive\ test\ reaction) = \frac{60\% \cdot 10\%}{60\% \cdot 10\% + 20\% \cdot 90\%} = \frac{6}{6+18} = 25\% \qquad (1)$$

This facilitating effect of natural frequencies is especially based on the evolutionary argument that human mind has been tuned to this kind of information format "long before the advent of probability format" (Gigerenzer & Hoffrage, 1995, p. 697). As a consequence, also the computation becomes easier (cf. Johnson & Tubau, 2015): When comparing both terms for the same ratio in equation (1) only three absolute numbers instead of six percentages must be processed in the term.

Whereas the facilitating effect of the natural frequency format is scientifically well researched, the facilitating effect of visualizations is more ambiguous. For this, we research on the question which kind of visualization is especially powerful in increasing people's performance in Bayesian reasoning tasks. In this paper, we investigate the effectiveness of two visualizations, i.e. the tree diagram and the unit square.

THEORETICAL BACKGROUND OF VISUALIZING BAYESIAN SITUATIONS

Duval (1999) stresses the meaning of representation and visualization in the field of understanding mathematics. Following Kaput (1987, p. 22) "…the root phenomena of mathematics learning and application are concerned with representation and symbolization because these are at the heart of the content of mathematics and are simultaneously at the heart of the cognitions." Especially regarding statistics and probability this becomes important, because the ability to read and to interpret graphical representations of data is a fundamental statistical idea for teaching statistics in school (Burrill & Biehler, 2011).

There are different aspects which are fundamental for coping adequately with Bayesian situations: From a purely practical point of view, the foremost concern is to understand the problem situation and to get the right solution. Addressing this aspect, we speak of "Bayesian reasoning" in a narrow sense (cf. Böcherer-Linder, Eichler, & Vogel, 2017, p. 2). From an educational point of view of learning mathematics, additionally the understanding of the mathematical structure of Bayesian problem situations becomes important (cf. Eichler & Vogel, 2015). Addressing this aspect, we speak of "flexible Bayesian reasoning" in a broader sense (Böcherer-Linder et al., 2017, p. 2). An important issue concerning the flexible Bayesian reasoning aspect is, for example, "to investigate the influence of variations of input parameters on the result" Borovcnik (2012, p. 21). Subsequently, the question of visualizing Bayesian situations can be considered regarding different objectives.

By mainly focusing on tree diagrams Spiegelhalter and Gage (2014) emphasize the need of visualizing Bayesian situations for facilitating a specific situation's interpretation and allowing people to estimate a risk adequately. It has been shown several times that trees (e.g. Fig. 3, left side) can support Bayesian reasoning (e.g. Sedlmeier & Gigerenzer, 2001; Wassner, 2004; Binder et al., 2015). Whereas there is a considerable amount of different visualizations for communicating Bayesian situations (Khan, Breslav, Glueck, & Hornbæk, 2015), research on visualizations of Bayesian situations for students' learning is mainly restricted firstly to the Bayesian reasoning, but not the flexible Bayesian reasoning mentioned above and secondly, to the tree diagram (as well as to 2x2-tables, cf. Veaux, Velleman, & Bock, 2011) but not to the unit square (Fig. 3, right side) which we plead for.

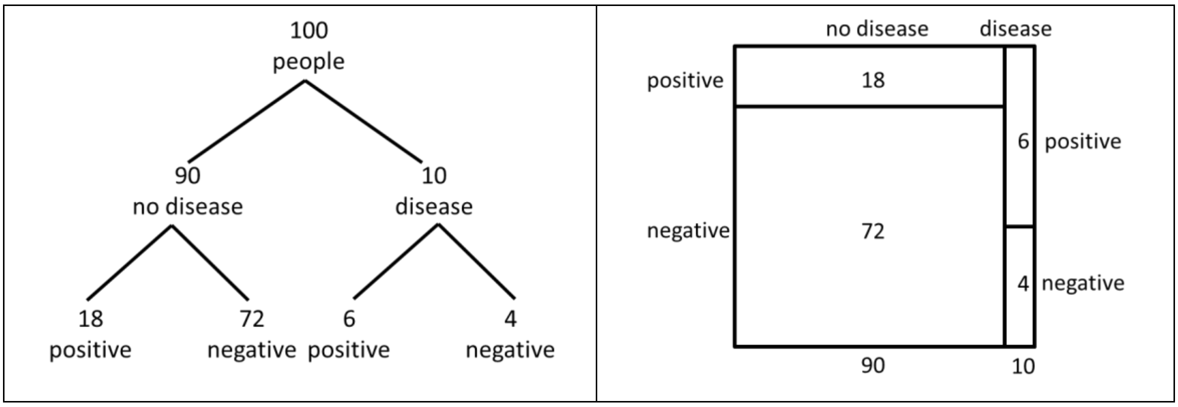

Figure 3. Medical diagnosis situation represented in the tree (left) and the unit square (right)

Referring to the diagnosis task represented in the frequency format (Fig. 2), the unit square is partitioned into four areas (Fig. 3) concerning the events having the disease (D), not having the disease ($\overline{D}$), getting a positive test result (T+) or a negative test result (T-). The vertical partitioning is determined by the event of having the disease which corresponds to the probability P(D) = 10% of having the disease and accordingly to the probability P($\overline{D}$) = 90% of not having the disease. The horizontal partitioning depends on the vertical partitioning and, thus, represent conditional events, which correspond to the probability that a healthy person gets wrongly a positive test result ($\mathrm{P}(\mathrm{T}+|\overline{D})$ (left side above) and accordingly to the probability that a person having the disease gets correctly a positive test result ($\mathrm{P}(\mathrm{T}+|\mathrm{D})$ (right side above). The areas represent joint probabilities, i.e.$\mathrm{P}(\mathrm{D} \cap \mathrm{T}+)$, $\mathrm{P}(\mathrm{D} \cap \mathrm{T}-)$, $\mathrm{P}(\overline{D} \cap \mathrm{T}+)$ and $\mathrm{P}(\overline{D} \cap \mathrm{T}-)$. The natural frequencies shown in figure 3 represent from the perspective of probability theory the expected values for the compounded events, i.e. 6 (D∩T+), 4 (D∩T-), 18 ($\overline{D}$ ∩T+) and 72 ($\overline{D}$ ∩T-).

The unit square is a statistical graph (Tufte, 2015), which means, that the sizes of the partitioned areas are proportional to the sizes of the represented data. Therefore, the proportions of incidences, like e.g. the base-rate, in a population are represented numerically as well as geometrically. From a theoretical point of view there are advantages of the visualization by the unit square concerning both the Bayesian reasoning as well as the flexible Bayesian reasoning.

- For finding the right solution of a typical Bayesian problem like in equation (1) the nominator as well as the denominator of the ratio representing the decisive subset relations are crucial. In the tree the relevant branches representing the ratio for the right solution are not directly related nor neighbored because they are not in line with the hierarchy of the tree diagram (Fig. 4, left side). On contrary, the relevant fields are always neighbored because the unit square has no hierarchical structure. Thus, we assume the unit square as being a more supportive visualization concerning Bayesian reasoning situations.
- Concerning the flexible Bayesian reasoning the unit square is suited to visualize the change of influencing parameters like the base-rate. If there is, for example, a base-rate of 30% of women at age 60 to be considered in the diagnosis task, the changing situation can be displayed by the unit square via a "picture-formula" (Eichler & Vogel, 2010) because of the geometric properties of the unit square (fig. 4, right side). Thus, the unit square allows for predicting the resulting probability as becoming bigger without calculating exact values whereas the tree does not.

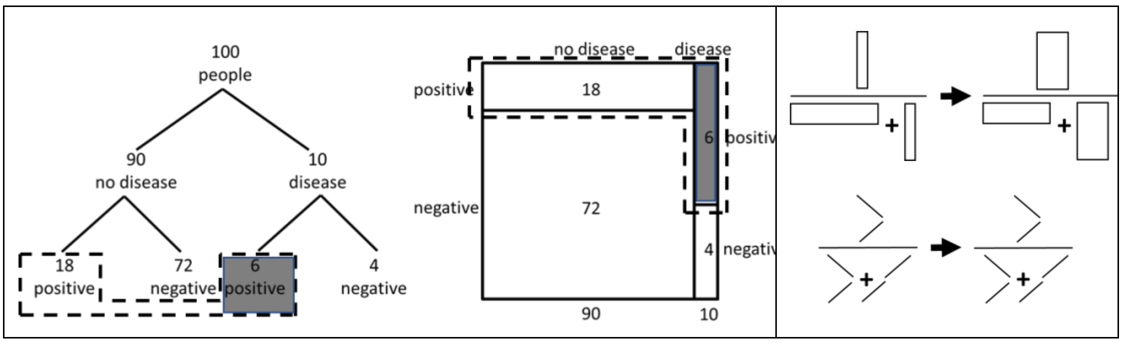

Figure 4. Tree and unit square in situations of Bayesian reasoning and flexible Bayesian reasoning

EMPIRICAL STUDIES

When asking for increasing people's performance in Bayesian reasoning an obvious first question is if the strategies of facilitating via the format of natural frequencies and via the format of visualizations can be combined. In fact, Garcia-Retamero and Hoffrage (2013) found a facilitating effect of visualizations in addition to the natural frequencies. However, it remains as an open question, which visualization is most efficient and which features lead to a facilitating effect (Binder et al., 2015). Beyond this, we are especially interested in an empirical answer to the question which visualization and furthermore, which specific features of these different visualizations are crucial for supporting people in Bayesian reasoning situations as well as in flexible Bayesian reasoning situations. Within our research group (Andreas Eichler and this paper's authors) three experiments were conducted with university students by reducing on trees and unit squares both based on format of natural frequencies to investigate these questions. In concrete, the tree diagram and the unit square were examined referring to their effect on students' computing set-subset relations in Bayesian situations as well as on their applying Bayes' rule. This aspect represents demands on Bayesian reasoning. Furthermore, the effect of a changed base rate representing demands on flexible Bayesian reasoning was part of the investigation. In the meanwhile, the results of these three experiments are published in Böcherer-Linder and Eichler (2017) and Böcherer-Linder et al. (2017). In the following, we firstly summarize the most important cornerstones of these experiments with university students. Thus, the arrangement of the fourth study which focused on students of the lower secondary level can be explained by being in logical succession with the preceding experiments.

Experiment 1

In the first experiment 148 teacher students beyond the first year enrolled in a course of mathematics education participated, neither having heard something about Bayes' rule nor the visualizations (tree, unit square) at the university before. In the test they were asked to compute different ratios among set-subset relations given in a Bayesian reasoning situation. In terms of the medical diagnosis situation they would have had to calculate, for example, the proportion of not infected and negatively tested people among all people or the proportion of the people having the disease among people being positively tested. This last proportion represents the structure of a typical Bayesian problem situation, which is, as mentioned above, more transparently visualized by the unit square than by the tree diagram (fig. 4, left side). The results provided significant empirical evidence for the unit square being more supportive especially referring those items which addressed demands of Bayesian reasoning.

Experiment 2

The participants of the second experiment were 143 undergraduates in the second year of their study enrolled in a course of Electrical Engineering. In comparison to the participants of the first experiment they must be judged as being more intensively and technically trained in mathematics at the university courses. However, concerning Bayes reasoning situations they also have not been schooled at the university before. In experiment 2 the participants were asked by four different items to apply Bayes' rule, like for example: "Calculate the ratio of people afflicted with the disease among the people tested positive". Therefore, four different Bayesian problem contexts were varied. The procedure was the same as in experiment 1. The results indicated a statistically significant effect in favor of the unit square, underlined by a large effect size.

Experiment 3

The third experiment followed immediately the second one to get information about those participants' ability of flexible Bayesian reasoning. Therefore, they got nine items that addressed the impact of a changed base rate. For example, one of these items was: "How is the change of the following ratio if the proportion of people that are infected with the disease would be bigger? The ratio of people tested negative among the infected people will be bigger / smaller / constant." The procedure was the same as in experiment 2. Based on theoretical considerations (cf. above) the unit square was expected to be more supportive because within this kind of visualization, for example, a changing base rate can be imaginable in front of the learners' inner eyes by moving a unit square's vertical line in the middle to the left or to the right. The nine test-items actually yielded empirical evidence for these theoretical assumptions.

By summarizing these three experiments being part of the ongoing research of our research group it can be stated: The results of the three experiments confirmed our hypotheses that the unit square is more efficient to facilitate Bayesian reasoning as well as flexible Bayesian reasoning referring the visibility of the relevant mathematical structure of the problem situation. Recognizing this structure is crucial when computing set-subset relations in Bayesian situations, applying Bayes' rule and estimating the effect of a changed base rate. A limitation of these studies must be seen in the level of the educational background of the participants. They all were students at university which reached university entrance level by having successfully finished the upper secondary school before.

Accordingly, we ask for the possibility of transferring the unit square's supporting effects into the level of lower secondary school. On this base, we conducted a replication study with students of lower secondary level to research on potential effects of mathematical education. In the following, we will report on the current state of this fourth experiment in more detail. The final results will be presented at the conference.

Experiment 4

Subjects: The participants are about 140 students of six classes on lower secondary level of different schools in the region Rhein-Neckar neighbored to Heidelberg in Germany. Since the tree diagram is part of the curriculum of the federal state of Baden-Württemberg, the students were expected to have some pre-knowledge concerning this kind of visualization but not within the context of (flexible) Bayesian reasoning problems which are not part of the curriculum. The unit square has to be considered as being unknown because of (still) not being contained in mathematical text books for German secondary schools.

Material and procedure: The participants were randomly assigned to either a unit square condition or a tree condition. The tasks in both conditions (unit squares and tree diagrams) were identical, only the visualizations differed (comparable to fig. 3). For purposes of replicating the forgoing studies in school the corresponding items were used again by asking for computing different ratios among set-subset relations given in Bayesian reasoning situations (cf. experiment 1), for applying Bayes' rule (cf. experiment 2), and for flexible Bayesian reasoning (cf. experiment 3). These items were used by varying different Bayesian problem contexts. These contexts were chosen regarding their familiarity of the students' daily live experience (for example, boys/girls wearing/not wearing bicycle helmets). Each subsample was provided with a brief description of the corresponding visualization. The test took 30 minutes. The participation was voluntary, and anonymity was guaranteed to the participants as well as to their parents.

Hypotheses: Our theoretical arguments in favor of the unit square were based on having additional geometrical features but not the viewers' pre-knowledge about Bayesian reasoning problems. Furthermore, the experiments 1 and 2 yielded some empirical evidence that the supportive function of the unit square holds across different groups of mathematical expertise concerning Bayesian reasoning tasks. Thus, we hypothesize the unit square as being more supportive for lower level secondary students when solving Bayesian reasoning problems. This effect will be expected as being larger in case of flexible Bayesian reasoning tasks because only the unit square with its geometrical properties allows for qualitatively calculating changing ratios whereas the tree diagram does not.

Results: The survey and the data-analysis are currently going on. Results of the fourth study will be presented at the conference.

CONCLUSION

Communication of risk and uncertainty are "hot topics" (Spiegelhalter & Gage, 2014) where the coping with (flexible) Bayesian reasoning problems plays an important role. The theoretical considerations and the four empirical studies of our research group reported on here intend to enrich the debate towards efficient features of visualizations of Bayesian situations to support people becoming more competent in decision making under uncertainty.

REFERENCES

Binder, K., Krauss, S., & Bruckmaier, G. (2015). Effects of visualizing statistical information – an empirical study on tree diagrams and 2 × 2 tables. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01186

Böcherer-Linder, K., & Eichler, A. (2017). The Impact of Visualizing Nested Sets. An Empirical Study on Tree Diagrams and Unit Squares. *Frontiers in Psychology*, *7*, 2026. https://doi.org/10.3389/fpsyg.2016.02026

Böcherer-Linder, K., Eichler, A., & Vogel, M. (2017). The impact of visualization on flexible Bayesian reasoning. *Avances de Investigación en Educación Matemática*. (11).

Borovcnik, M. (2012). Multiple perspectives on the concept of conditional probability. *AIEM. Avances de Investigación en Educación Matemática*, *1*, 5–27.

Burrill, G., & Biehler, R. (2011). Fundamental Statistical Ideas in the School Curriculum and in Training Teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *New ICMI Study Series: vol. 14. Teaching statistics in school mathematics, challenges for teaching and teacher education: A joint ICMI/IASE study: the 18th ICMI study*. Dordrecht: Springer.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, *58*(1), 1–73. https://doi.org/10.1016/0010-0277(95)00664-8

Duval, R. (1999). Representation, Vision and Visualization: Cognitive Functions in Mathematical Thinking. Basic Issues for Learning. Retrieved from https://files.eric.ed.gov/fulltext/ED466379.pdf

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge, New York: Cambridge University Press.

Eichler, A., & Vogel, M. (2010). Die (Bild-)Formel von Bayes [The (picture-)formula of Bayes]. *PM - Praxis der Mathematik in der Schule*, *52*(32), 25–30.

Eichler, A., & Vogel, M. (2015). Teaching Risk in School. *The Mathematics Enthusiast*, *12*(1), 168–183. Retrieved from http://scholarworks.umt.edu/tme/vol12/iss1/17

Garcia-Retamero, R., & Hoffrage, U. (2013). Visual representation of statistical information improves diagnostic inferences in doctors and their patients. *Social Science & Medicine, 83*, 27–33.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve bayesian reasoning without instruction: Frequency formats. *Psychological review*, *102*(4), 684–704.

Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic medicine : journal of the Association of American Medical Colleges*, *73*(5). Retrieved from http://library.mpib-berlin.mpg.de/ft/gg/GG_Using_1998.pdf

Johnson, E. D., & Tubau, E. (2015). Comprehension and computation in Bayesian problem solving. *Frontiers in Psychology*, *6*, 938. https://doi.org/10.3389/fpsyg.2015.00938

Kaput, J. J. (1987). Representation Systems and Mathematics. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 19–26). Hillsdale, NJ: Lawrence Erlbaum Associates.

Khan, A., Breslav, S., Glueck, M., & Hornbæk, K. (2015). Benefits of visualization in the Mammography Problem. *International Journal of Human-Computer Studies*, *83*, 94–113. https://doi.org/10.1016/j.ijhcs.2015.07.001

McDowell, M., & Jacobs, P. (2017). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin*, *143*(12), 1273–1312. https://doi.org/10.1037/bul0000126

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology: General*, *130*(3), 380–400. https://doi.org/10.1037//0096-3445.130.3.380

Spiegelhalter, D., & Gage, J. (2014). What can we learn from real-world communication of risk and uncertainty? In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education.: Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.*. Voorburg, The Netherlands: International Statistical Institute (ISI).

Tufte, E. R. (2015). *The visual display of quantitative information* (2. ed., 9. printing). Cheshire, Conn.: Graphics Press.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131.

Veaux, R. D. de, Velleman, P. F., & Bock, D. E. (2011). *Intro stats*. Boston: Addison Wesley.

Wassner, C. (2004). *Förderung Bayesianischen Denkens: Kognitionspsychologische Grundlagen und didaktische Analysen*. *Texte zur mathematischen Forschung und Lehre: Vol. 34*. Hildesheim: Franzbecker.