

## CATEGORIZING ERRORS IN BAYESIAN SITUATIONS

Andreas Eichler and Katharina Böcherer-Linder

Institute for Mathematics, University of Kassel, Germany

Institute for Mathematics Education, University of Education Freiburg, Germany

eichler@mathematik.uni-kassel.de

*Many researchers mentioned that learners struggle when dealing with Bayesian situations, i.e. in situations that require using Bayes' formula. It is further a wide consensus that both natural frequencies and visualizations facilitate learners' performance in Bayesian situations. In this contribution we refer to five visualizations of Bayesian situations, i.e. a tree diagram, a double-tree, an icon array, a unit square and a 2x2-table and investigate the errors that learners show when solve problems in Bayesian situations. We categorize these errors and link the errors to the specific characteristics of the visualizations.*

### BACKGROUND

Bayesian situations, i.e. situations that require using Bayes' formula, are known as a subject in the field of probability that seems to be against human intuition (Tversky & Kahneman, 1974). Many researchers in probability education or cognitive psychology described errors that professionals and layman make when dealing with Bayes' formula (e.g. Diaz & Batanero, 2009; Eddy, 1982; Zhu & Gigerenzer, 2006). A meta-analysis of McDowell and Jacobs (2017) yielded that only about 5% of participants in different studies were able to solve a task in a Bayesian situation like given in Figure 1 that shows an often used example of a Bayesian situation representing a medical diagnosis test (c.f. Johnson & Tubau, 2015).

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.  
What is the probability of having the disease given that the test is positive?

Solution:  $P(\text{disease}|\text{test positive}) = \frac{P(\text{disease and test positive})}{P(\text{test positive})}$

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

However, research in recent two decades yield two strategies that seem to facilitate dealing with Bayesian situations. The first strategy is to represent the statistical information of a Bayesian situation with natural frequencies (Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2000). That means to use countable entities as a fictive population and to develop the sampling process based on this fictive population (Hoffrage, Gigerenzer, Krauss, & Martignon, 2002). The definition of a fictive population of 100 women and the translation of the Bayesian situation given above in the format of natural frequencies is shown in Figure 2.

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.  
What is the proprtion of having the disease given that the test is positive?

Solution:  $P(\text{disease}|\text{test positive}) = \frac{P(\text{disease and test positive})}{P(\text{test positive})} = \frac{6}{24} = 0,25$

Figure 2. Medical diagnosis test in the format of natural frequencies.

The meta-analysis of McDowell and Jacobs (2017) yielded that using the format of natural frequencies increases the performance of participants in different studies from about 5% to about 25%. A second strategy is to use visualizations of a Bayesian situation (e.g. Brase, 2009). Although studies using visualizations in addition to natural frequencies reported an increase of the rate of

correct solution in different Bayesian situations to 40% to 70% (Binder, Krauss, & Bruckmaier, 2015; Böcherer-Linder & Eichler, 2017), however, the facilitating effect of visualization is not as clear as the facilitating effect of natural frequencies (Johnson & Tubau, 2015; McDowell & Jacobs, 2017).

One reason of the ambiguity of results referring the strategy of visualizations could be identified in different styles of visualizations, i.e. a branch style, a nested style and a frequency style (Khan, Breslav, Glueck, & Hornbæk, 2015). The branch style, e.g. a tree diagram (Fig. 3), represents sets and subsets in a Bayesian situation in a hierarchical structure. The nested style, e.g. the unit square, comprise sets and subsets in a Bayesian situation as neighbored fields (Fig. 3) and, finally, the frequency style shows in addition countable entities representing the statistical information (Fig. 3).

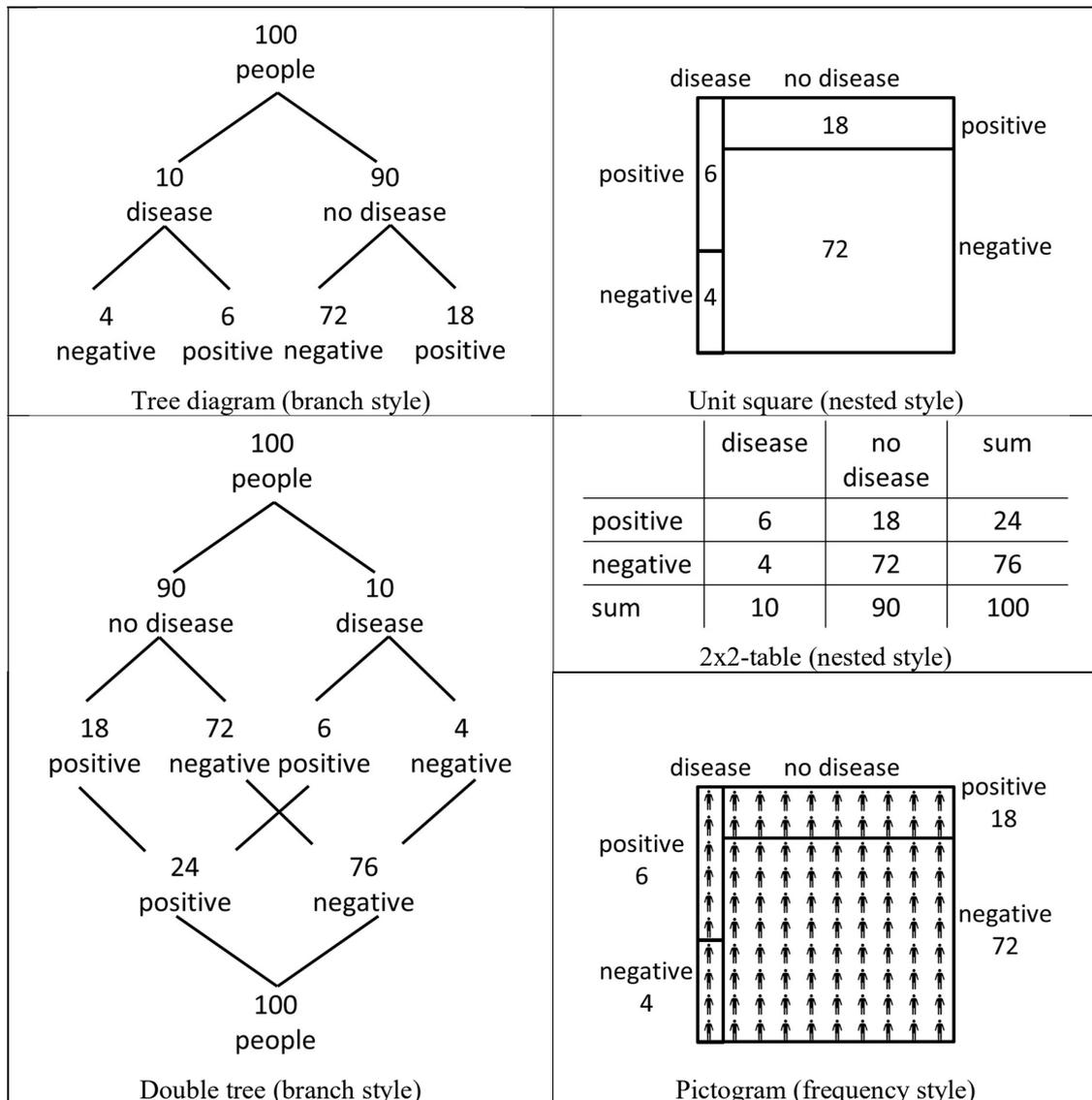


Figure 3. Five visualizations of the medical diagnosis test

To investigate further the additional facilitating effect of visualizing Bayesian situations in which the statistical information is represented by natural frequencies, three directions of research are reasonable.

A first direction involves an investigation of performance rates of different visualizations like shown in figure 3. On the basis of using natural frequencies, for example, the research of

Binder et al. (2015) implies that a 2 x 2-table is more effective than the tree diagram. Further, our own research gave evidence that the unit square is significantly more effective than the tree diagram (Böcherer-Linder & Eichler, 2017).

A second direction could focus on properties of different visualizations that could theoretically explain different performance rates. For example, in our own research mentioned above, we showed that the visualization of the tree diagram and the unit square differed considerably in making transparent the set-subset relation that is used in Bayes formula: While the set of positive tested women is visualized by neighboring fields in the unit square, this set is not visualized by one node in the tree diagram but only in two branches that are not neighbored.

In this paper we focus on the third direction to investigate differences in the facilitating effect of visualizations, i.e. to analyze the solution of those who make errors in dealing with Bayesian situations. For this, we refer to existing research concerning the categorization of errors in Bayesian situations. We further exemplarily analyze the five visualizations illustrated above from the perspective of errors when dealing with Bayesian situations. Finally, we describe the method of our study to investigate university students' errors in different Bayesian situations. Since we are not able to present the results of our ongoing study in this paper, we will present these results at ICOTS10.

CATEGORIES OF ERRORS WHEN DEALING WITH BAYESIAN SITUATIONS

For regarding errors when dealing with Bayesian situations we exemplarily use two of the five visualizations representing different styles, i.e. the tree diagram and the unit square (Fig. 4). In these visualizations we indicate the sets and subsets or rather the cardinal number of these sets (natural frequencies) with letters. We define the quotient  $\frac{C_2}{C_2+C_4}$  as the solution in a Bayesian situation using Bayes formula. Actually, also other quotients, e.g.  $\frac{C_1}{C_1+C_3}$ , could represent a solution in a Bayesian situation, but every of these quotients could be identified with the given solution when the numbering of sets and subsets would be appropriately changed.

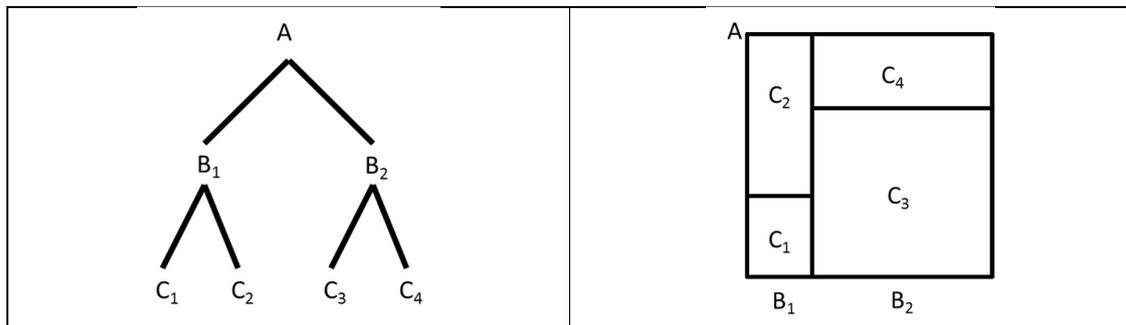


Figure 4. Tree diagram and unit square with indication of sets and subsets

Zhu and Gigerenzer (2006) gave in two studies about 200 students of different ages seven and, respectively, ten Bayesian situations, where the statistical data was represented by natural frequencies. They found four categories of errors that a substantial number of participants made.

The most significant error labeled as non-Bayesian strategy was a strategy that Zhu and Gigerenzer (2006) called “pre-Bayes” and that is represented by the quotient  $\frac{B_1}{C_2+C_4}$ . They found this strategy in about 10% of all solutions. Based on the medical diagnosis task shown in Fig. 2, these participants had to find the proportion of people having the disease given a positive test. However, they used inappropriately the entire set of people having the disease instead of the intersection of the set of people having the disease *and* the set of positively tested people for their solution.

Further, about 5% of the participants made an error that Zhu and Gigerenzer (2006) called “conservatism” represented by the quotient  $\frac{B_1}{A}$ . This error was also reported as “base-rate only” by Gigerenzer and Hoffrage (1995). In this case, the information of having a *positive test* is not regarded.

About 3% of the participants made an error that Zhu and Gigerenzer (2006) called “evidence-only” represented by the quotient  $\frac{C_2+C_4}{A}$ . This error was understood as opposite of the conservatism since in this case the first information, i.e. *having the disease*, is not regarded.

Only about 1,5% of the participants made an error that Zhu and Gigerenzer (2006) called “representative thinking” represented by the quotient  $\frac{C_2}{B_1}$ . This strategy was also reported by Dawes (1986) and Gigerenzer and Hoffrage (1995), who reported that nearly 6% of the participants used this strategy. In a broader context this strategy is also reported as the confusion between  $P(A|B)$  and  $P(B|A)$  when dealing with conditional probabilities (e.g. Diaz & Batanero, 2009).

An error type that was not found in Zhu and Gigerenzer (2006), but was significant in Gigerenzer and Hoffrage (1995) was called “joint occurrence strategy” and is represented by the quotient  $\frac{C_2}{A}$ . In this case, participants seem to over-emphasize the event of having the disease *and* getting a positive test result.

#### CONNECTIONS BETWEEN ERRORS AND PROPERTIES OF A VISUALIZATION

Based on existing research results of several studies (see the introduction) that yield different performances of participants when using different visualizations of Bayesian situations, it is a crucial question, if different visualizations also result in different distributions of errors. To investigate the distribution of errors based on different visualization is the main aim of this study.

There are two main differences between the two visualizations shown in Figure 4, i.e. the tree diagram and the unit square. Firstly, the two subsets and their cardinal numbers that have to be identified for the denominator in the correct solution, i.e.  $C_2$  and  $C_4$ , are represented by neighbored fields in the unit square but by two branches in the tree diagram that were not neighbored. This was the main issue of explaining significant differences in performance rates in Bayesian situations (c.f. Böcherer-Linder & Eichler, 2017). For this reason, a hypothesis is that students, who use the unit square, more frequently refer to the sum of  $C_2 + C_4$ , that is the denominator in the correct solution and also in the error strategies of “pre-Bayes” and “evidence-only”. It is further a hypothesis that the performances of participants using the unit square are not different from the performances of participants using the 2x2-table, the double tree or the icon array since all these visualizations has  $C_2$  and  $C_4$  as neighbored fields (2x2-table; icon array) or as a node (double tree).

Further, the set or rather its cardinal number  $B_1$  is represented by a node in the tree diagrams but not directly in the unit square and the icon array as shown in Figure 3. For this reason, it is a hypothesis that students using the tree diagrams more frequently refer to  $B_1$ .

With respect to the categories of errors that were reported in recent research the hypotheses mentioned above have to be modified as follows:

- The category “joint occurrence strategy” ( $\frac{C_2}{A}$ ) is not based on the use of  $B_1$  or  $C_2 + C_4$ . Thus, there seems to be no reason to expect differences in the solutions of participants using different visualizations.
- Although the category “representative thinking” is based on the use of  $B_1$ , also the shape of the unit square (the icon array and the 2x2-table) could trigger this error, since one could develop a quotient by focusing on one side of the diagram that is divided by the other side by a continuous line. For this reason, we do not expect differences when using the different visualizations.
- The category “evidence-only” is based on the use of  $C_2 + C_4$ . For this reason, we expect more solutions in this category for those participants who use the unit square, the icon array or the 2x2-table.
- The category “conservatism” is based on the use of  $B_1$ . For this reason, we expect more solutions in this category for those participants who use the tree diagram and also the double tree.
- The category “pre-Bayes” is based on the use of both  $B_1$  and  $C_2 + C_4$ . However, to take into account at least implicitly  $C_1$  as a part of  $B_1$  seems not to be plausible for those participants that use the unit square or the icon array (and also the 2x2-table). However, since  $B_1$  is a node in the

tree diagrams, we expect that more participants using tree diagrams choose the “pre-Bayes” strategy.

METHOD

To investigate the strategies of students in Bayesian situations, we defined six different situations. In every situation, the context and the population was roughly described. The statistical information was only given in the visualization (for an example see the visualization with the tree diagram for one Bayesian situation in Figure 5). The other situations are given in Table 1 using the notation for sets or rather cardinal numbers as given in Figure 4.

Economics (cf. Binder et al., 2015)	$A$ : 1000 university students $B_1$ : 325 students attend in an economics course; $B_2$ : 675 attend not $C_1, C_3$ : students are not career oriented; $C_2, C_4$ : students are career oriented Task: Calculate the proportion of career oriented students among the students that attend an economic course.
Clothes	$A$ : 100 clothes $B_1$ : 40 trousers; $B_2$ : 60 pullovers $C_1, C_3$ : red; $C_2, C_4$ : blue Task: Calculate the proportion of trousers among the blue clothes.
Medical diagnosis test	$A$ : 1000 people $B_1$ : 100 disease; $B_2$ : 900 no disease $C_1, C_3$ : test negative; $C_2, C_4$ : test positive Task: Calculate the proportion of people having the disease among the positively tested people.
Flowers	$A$ : 120 flowers $B_1$ : 50 carnations; $B_2$ : 70 roses $C_1, C_3$ : white; $C_2, C_4$ : red Task: Calculate the proportion of carnations among the red flowers.
Bulbs	$A$ : 100 bulbs $B_1$ : 40 sort A; $B_2$ : 60 sort B $C_1, C_3$ : not flourishing; $C_2, C_4$ : flourishing Task: Calculate the proportion bulbs of sort A among the flourishing bulbs.

Table 1: Five situations requiring using Bayes’ formula

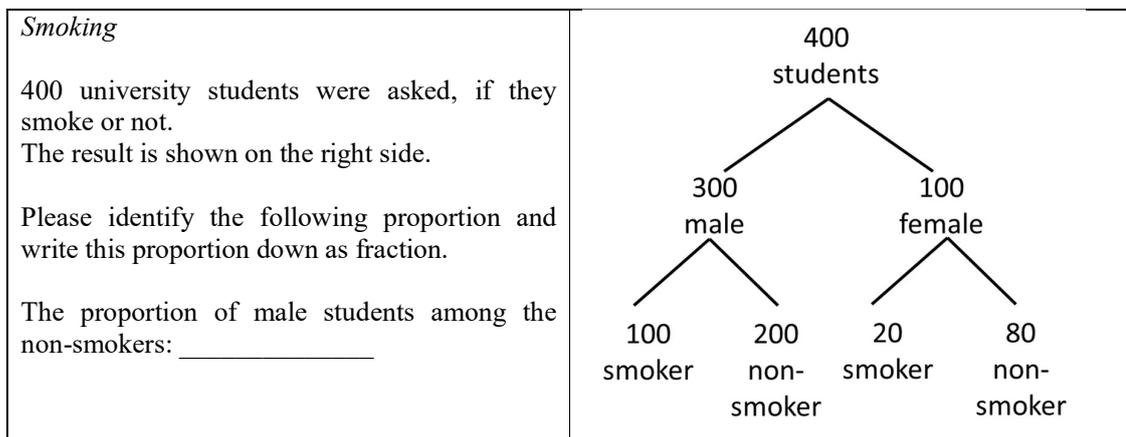


Figure 5. Description of one Bayesian situation with the tree diagram

The frequencies in every situation were developed in a way that allows to differentiate among the categories of errors described above.

We conducted the test with a sample of 276 undergraduate students that were assigned to one of the five visualizations by random. For every condition, there was on one page a brief description of the visualization.

## CONCLUSION

The basis for our analysis of errors that students made when dealing with Bayesian situations were 1535 completed solutions. At ICOTS10 we will have finished the analysis of errors and the analysis of the distribution of errors on the five visualizations described in this paper. A first impression is that particularly the tree diagram yield a distribution of errors that is considerably different to the error distribution of the unit square, the icon array and also the 2x2-table.

## REFERENCES

- Binder, K., Krauss, S., & Bruckmaier, G. (2015). Effects of visualizing statistical information - an empirical study on tree diagrams and  $2 \times 2$  tables. *Frontiers in psychology*, *6*, 1186. <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*(1186), 241. <https://doi.org/10.3389/fpsyg.2016.02026>
- Brase, G. L. (2009). Pictorial representations in statistical reasoning. *Applied Cognitive Psychology*, *23*(3), 369–381. <https://doi.org/10.1002/acp.1460>
- Dawes, R. (1986). Representative thinking in clinical judgment. *Clinical Psychology Review*, *6*(5), 425–441. [https://doi.org/10.1016/0272-7358\(86\)90030-9](https://doi.org/10.1016/0272-7358(86)90030-9)
- Diaz, C., & Batanero, C. (2009). University Students' Knowledge and Biases in Conditional Probability Reasoning. *International Electronical Journal of Mathematics Education*, *4*(3), 131–162.
- 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: Cambridge University Press.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684–704. <https://doi.org/10.1037/0033-295X.102.4.684>
- Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: What natural frequencies are and what they are not. *Cognition*, *84*(3), 343–352. [https://doi.org/10.1016/S0010-0277\(02\)00050-1](https://doi.org/10.1016/S0010-0277(02)00050-1)
- Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating Statistical Information. *Science*, *290*(5500), 2261–2262. <https://doi.org/10.1126/science.290.5500.2261>
- 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>
- 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*. Advance online publication. <https://doi.org/10.1037/bul0000126>
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science (New York, N.Y.)*, *185*(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: the role of representation in mental computation. *Cognition*, *98*(3), 287–308. <https://doi.org/10.1016/j.cognition.2004.12.003>