Analysing students' notes when calculating in Bayesian situations

Theresa Büchter, Nicole Steib, Karin Binder, Katharina Böcherer-Linder, Andreas Eichler, Stefan Krauss, Markus Vogel

Analysing students' notes when calculating in Bayesian situations

Theresa Büchter[1], Nicole Steib[2], Karin Binder[3], Katharina Böcherer-Linder[4], Andreas Eichler[1], Stefan Krauss[2] and Markus Vogel[5]

[1]University of Kassel, Institute of Mathematics, Germany; tbuechter@mathematik.uni-kassel.de

[2]University of Regensburg, Faculty of Mathematics, Germany

[3]Ludwig Maximilian University Munich, Institute of Mathematics, Germany

[4]University of Freiburg, Department of Mathematics Education, Germany

[5]University of Education, Heidelberg, Institute of Mathematics, Germany

Calculation in a so-called Bayesian situation is known to be challenging. According to a meta-analysis the performance is only 4% if the relevant information is given in form of probabilities. Still, previous research has identified helpful strategies for improving Bayesian reasoning (i.e., so-called natural frequencies and visualizations). Yet, it has hardly been studied, if and how students use these strategies themselves. Thus, in this paper we focus on the process instead of the result of students who work on a Bayesian situation. For that, we present a coding system for analysing the notes of the students and the strategies which they apply for the calculation. We identify that the the students hardly ever use the strategy of natural frequencies. Further, about half of the students use common visualizations such as a probability tree-diagram or a 2×2 table. Yet, completing the 2×2 table was error-prone as students often entered conditional (instead of joint) probabilities into the inner fields.

Keywords: Bayesian reasoning, visualization, natural frequencies.

Introduction

Imagine a person tests positive in a SARS-CoV-2 self-test, who just returned from a high-incidence area with symptoms of a cold. The sensitivity (96%) and the false-positive rate (2%) of the test are provided in the patient information leaflet and from current statistics it can be assumed that the probability to be infected is 5% for people returning from a high-incidence with symptoms of a cold. How would you determine the probability that this person (with a positive test result, symptoms of a cold and returning from a high-incidence area) is *actually* infected? Different approaches to calculate the requested probability are possible, e.g., using Bayes' formula, or translating the given information into a fictitious sample with expected frequencies and calculating the correct proportion of the relevant frequencies. Further, drawing a visualization such as a tree-diagram could support the identification of the correct algorithm. From previous research, some of these strategies seem more or less promising for working on a calculation in a so-called *Bayesian situation*. The described scenario is an example of such a Bayesian situation (Zhu & Gigerenzer, 2006), as an indicator (e.g., the positive SARS-CoV2 self-test result) is used to make inferences about a hypothesis (e.g., infection of the person). Previous research about Bayesian reasoning (i.e., the ability to adequately reason in a Bayesian situation) has shown that laymen and experts alike struggle immensely to reason in a Bayesian manner (e.g., McDowell & Jacobs, 2017). However, helpful strategies for the *presentation* of Bayesian situations have been identified to improve Bayesian reasoning, i.e., using natural frequencies and visualizations. Yet, these strategies have hardly been studied in a process-oriented approach which analyzes *how* participants work on problems in Bayesian situations and if they use these strategies themselves (for exceptions see e.g., Weber et al., 2018). Therefore, in this contribution it is analyzed if and how these previously identified helpful strategies are used by students without specific instruction. Therefore, first, the empirically identified helpful strategies are described, second, the methodology of the study and the coding system are characterized, with which students' notes were analyzed and third, the results of the strategies, which are used spontaneously, are presented and discussed afterwards. The participants of our study are medical and law students for whom Bayesian reasoning is of particular importance professionally (e.g., Hoffrage et al., 2000).

Theoretical Background

Bayesian reasoning is the ability to evaluate Bayesian situations, in which an indicator (I) is used to make inferences on a hypothesis (H). Usually, in a Bayesian situation three parameters are given (McDowell & Jacobs, 2017), the so-called base rate, $P(H)$, sensitivity, $P(I|H)$ and false-positive rate, $P(I|\overline{H})$. In previous studies, Bayesian reasoning has most often been measured by calculating a conditional probability such as the so-called positive predictive value (PPV), $P(H|I)$, when the prior three parameters are given. For that, Bayes' rule can be used to determine the positive predictive value: $P(H|I) = \frac{P(H)\cdot P(I|H)}{P(H)\cdot P(I|H)+P(\overline{H})\cdot P(I|\overline{H})}$. The ability to calculate a conditional probability, is assumed to be only one aspect of Bayesian reasoning, which we call *calculation*. Other aspects of Bayesian reasoning entail *covariation*, i.e., the ability to assess influences of changes in the given parameters on the PPV, and *communication*, i.e., the ability to adequately communicate probabilities in a Bayesian situation (Büchter et al., 2022). In this paper the focus is on the aspect of calculation.

Previous research on Bayesian reasoning has shown that performance of calculation is generally very poor (McDowell & Jacobs, 2017). Yet, two strategies can increase performance of calculation:

1. *Natural frequencies* as the format of the statistical information
2. *Visualizations* as a tool to structure the Bayesian situation

The first helpful strategy is the use of so-called natural frequencies for presenting the given statistical information, in which pairs of natural numbers represent the expected frequency in a fictious sample (Gigerenzer & Hoffrage, 1995). In Table 1 the statistical information of the Bayesian situation about SARS-CoV-2 antigen tests is presented in probabilities as well as in natural frequencies. It is observable that the calculation of the positive predictive value is computationally simpler when the information is given in form of natural frequencies. A meta-study shows that the performance increases to 24% if the statistical information is given in form of natural frequencies compared to 4% if the statistical information is given in probabilities (McDowell & Jacobs, 2017). Consequently, a first strategy which might be employed for working on the calculation aspect might be to translate the given probability information into natural frequencies, which is called the *frequency strategy*.

The second helpful strategy concerns the presentation of the structure of the Bayesian situation in a visualization. Depending on the specific visualization more and less beneficial effects have been documented. For instance, visualizations generally seem to be more supportive, if their numerical information is given in natural frequencies instead of probabilities (e.g., Binder et al., 2015; Binder et al., 2020). Moreover, Euler diagrams did not prove to be very supportive (Micallef et al., 2012) and the well-known tree-diagram is less likely to increase performance than other visualizations (Böcherer-Linder & Eichler, 2019), such as the so-called double tree, frequency net and unit square. Using these visualizations with natural frequencies, increases the performance to 50%-60%, thus also adds a significant facilitating effect to natural frequencies only (Binder et al., 2020; Böcherer-Linder & Eichler, 2019). However, they are not wide-spread and therefore not very well-known visualizations. Using a visualization as a strategy for working on the calculation aspect of the Bayesian situation is called the *visualization strategy* in the following.

Table 1: Bayesian situation about SARS-CoV-2 tests in probabilities and natural frequencies

	Probabilities	**Natural frequencies**
Context	Imagine, you are working as a general practitioner in your own doctor‘s office. You are currently consulting a patient with symptoms of a cold who has returned from a high-incidence area and used the „AESKU.RAPID“ self-test in order to detect an infection with SARS-CoV-2. The following statistics are known about the situation and AESKU.RAPID self-test:	
Statistical information	• The probability is **5%** that a person is infected with SARS-CoV-2. • If a person is infected with SARS-CoV-2, then the probability is **96%** that the person tests positive. • If a person is <u>not</u> infected with SARS-CoV-2, then the probability is **2%** that the person tests positive nevertheless.	• **50 out of 1000** people are infected with SARS-CoV-2. • **48 out of the 50** people, who are infected with SARS-CoV-2, test positive. • **19 out of the 950** people, who are <u>not</u> infected with SARS-CoV-2, test positive nevertheless.
Question about PPV	How likely is the patient actually infected with SARS-CoV-2, given the positive test result?	
Solution	$\frac{0.05 \cdot 0.96}{0.05 \cdot 0.96 + 0.95 \cdot 0.02} \approx 71\%$	$\frac{48}{48+19} \approx 71\%$

Previous research indicates that the specific design of a visualization further supports Bayesian reasoning (e.g., Clinton et al., 2016), e.g., by highlighting the relevant information in the visualization (compare also color-coding in research on multimedia-learning). Therefore, emphasizing the relevant information in a visualization could be an additional strategy, which we call the *highlighting strategy*.

Finally, the completeness of numerical information about the different sets and subsets in the Bayesian situation is assumed to influence the performance of calculation (Binder et al., 2022). For instance, Khan et al. (2015) have found that performance was comparable between a group of participants which received information about the Bayesian situation in a double tree vs. in a text, if the numerical information was equivalent in both groups. Therefore, a final strategy for working on the aspect of calculation in Bayesian situations may be to determine all subsets, which we call the *completing strategy* in the following.

Research interest

In contrast to previous research, the focus of this study is not on the performance of calculation but on the process-oriented aspects of calculation in a Bayesian situation. The particular interest is an exploratory analysis on the helpful strategies which are used by the participants in the current study.

Methodology

Participants

As Bayesian reasoning is professionally required among practitioners in the legal as well as medical domain (Hoffrage et al., 2000), participants of our study are students of law (*N*=255) and medicine (*N*=260) who worked on an authentic context within their domain. Like that, we hoped to ensure that the participants were motivated to thoroughly work on the task. Both fields of study do not contain obligatory courses in stochastics. However, calculation in Bayesian situations may be included in the medical state exam and therefore, self-studying among the medical students is possible.

Study design

The results are part of a larger study in which the participants received different training courses on Bayesian reasoning and the learning effect was measured in a pre-post-follow-up-design (http://bayesianreasoning.de/en/br_trainbayes_en.html). For the analyses described here, the focus is on the strategies used *without* any training and therefore on the pretest in one of three contexts of the pretest. The results represent the strategies used either in a medical context about the SARS-CoV-2 tests (Table 1) for the medical students or about a legal context about breath analyzer tests for the law students (compare Büchter et al., 2022). In both contexts base rate, sensitivity and false-positive rate were given in probabilities (without any visualization) and the students were asked to calculate the PPV. The Bayesian situation was presented digitally, but the participants had a blank sheet of paper and pen and colored pencils to take notes.

Coding system for analyzing the students' notes

A coding system has been developed according to a qualitative content analyses with deductively derived categories (Mayring, 2015). Each of the four strategies explained in the theoretical background (visualization, frequency, highlighting and completion strategy) forms a separate deductive category in the coding system, for which multiple dichotomous questions (0=no, 1=yes) cover the different manifestations (compare Table 2). The coding system is used for notes in the whole study and thus questions have been added, which are more relevant after participation in the training courses (such as question 2 in the highlighting strategy or questions 4 to 9 in the visualization strategy as well as questions 5, 6 and 8 about the completion strategy which are only relevant in tasks other than the presented one, which require calculation of the so-called negative predictive value) and are therefore printed in grey in Table 2. If multiple visualizations are used, then all other categories are coded separately for each of the visualizations (e.g., it is coded, which of the pieces of information are listed and which are correct in each of the visualizations). Initially, the notes of 25 of the 515 participants have been coded based on the four deductive categories. Thereby, observations have been made, which were not yet covered with the deductive categories. Thus, additional questions have been added inductively in order to identify participants who have in principle applied a conceptually correct approach but made a less crucial mistake in the process (questions 1 and 2) or who had simply memorized Bayes' formula and applied it without using any of the helpful strategies (question 3). After refining the coding system, the notes of the initially coded 25 participants have been double-coded and inter-rater reliabilities were measured with Cohens Kappa coefficients and are above .84 for all questions and consequently show a substantial consensus between different raters.

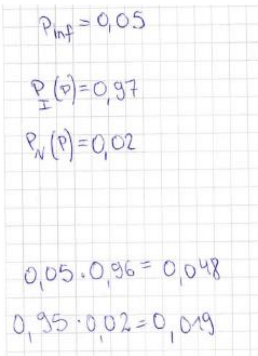
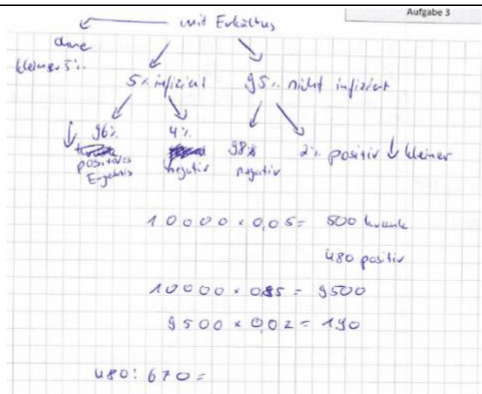
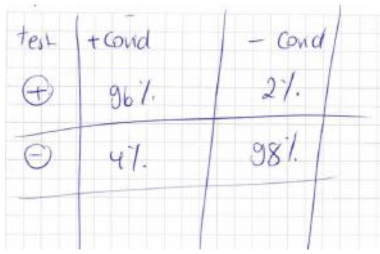
Table 2: Coding system for analyzing the notes from working on a calculation in a Bayesian situation

Category	Questions in the coding system
Visual representation of the Bayesian situation (*Visualization strategy*)	Is the Bayesian situation visualized… 1. … in a 2×2 table? 2. … in a tree-diagram? 3. … in a structure “reminiscent” of a tree-diagram? 4. … in a double-tree? 5. … in a structure “reminiscent” of a double-tree? 6. … in a unit square? 7. … in a structure “reminiscent” of a unit square? 8. … in a structure of frequencies? 9. … in a structure “reminiscent” of a structure of frequencies? 10. … differently?
Representation of the numerical information (*Frequency strategy*)	Do the notes contain… 1. … absolute frequencies? a. Does the sum of the frequencies equal 100? 2. … percentages? 3. … decimal fractions? 4. … common fractions?
Highlighted elements in the visual representation (*Highlighting strategy*)	Are parts of the notes highlighted… 1. … in general? 2. … in the same way as in the training courses?
Complete (& correct) representation of all probabilities or quantities in the Bayesian situation (*Completion strategy*)	Which of the following information in the notes are *a) listed* and *b) correct*? 1. Given information: $P(H), P(I\|H)$ and $P(I\|\overline{H})$ 2. Base rate information as P(H) and $P\left(\overline{H}\right)$ or as the quantities of H and $\overline{H}$ 3. Information about true-positives as $P(H \cap I)$ or as the quantity of $H \cap I$ 4. Information about false-positives as $P(\overline{H} \cap I)$ or as the quantity of $\overline{H} \cap I$ 5. Information about false-negatives as $P(H \cap \overline{I})$ or as the quantity of $H \cap \overline{I}$ 6. Information about true-negatives as $P(\overline{H} \cap \overline{I})$ or as the quantity of $\overline{H} \cap \overline{I}$ 7. Information about positives as $P(I)$ or as the quantity of I 8. Information about negatives as $P(\overline{I})$ or as the quantity of $\overline{I}$ 9. All pieces of information, which are used in the training *(only asking whether they are all listed)* 10. Are there calculations for any of the relevant probabilities or quantities without displaying the result?
Inductively derived characteristics of the notes	1. Are there hasty mistakes in any of the pieces of information, e.g., a mistake in copying the given information or in changing between different numerical representations (i.e., from percentages to fractions)? 2. Is the correct fraction for the requested probability in the notes? 3. Is Bayes’ formula in the notes? 4. Are the proportions of the areas in the unit square roughly adequate (only applicable for the unit square)? 5. Are the proportions of the areas in the unit square very precise, i.e., true to scale (only applicable for the unit square)?

Results

Different examples for notes of the students are displayed in Table 3. These three examples of the notes of participants are chosen to illustrate the strategies used most often by the participants.

Table 3: Examples of notes by three participants from working on calculation in a Bayesian situation

A) No visualization	B) Probability tree-diagram	C) Probability 2×2 table
		

In example B) it can be seen, that on the one hand a tree-diagram was used to visualize the given information (question 2, visualization strategy) and additionally, separate notes were added beneath the visualization (coded separately). In the tree diagram, only percentages are used (question 2, frequency strategy), while frequencies and decimal fractions are used in the notes beneath the tree diagram (questions 1 and 3, frequency strategy). Nothing is highlighted in either the tree-diagram or the notes. The tree-diagram contains the given information, as well as the base-rate information and both is correct (questions 1a, 1b, 2a and 2b, completion strategy). In the notes the given information (now represented as decimal fractions) are translated to frequencies with a sample of 10,000 including the base rate information (500 and 9500), the true-positives (480) and the false-positives (190) and all of these are correct (questions 1a, 1b, 2a, 2b, 3a, 3b, 4a, and 4b, completion strategy). Examples A), B) and C) are illustrative of the strategies observed in the pretest, as can be seen in Figure 1.

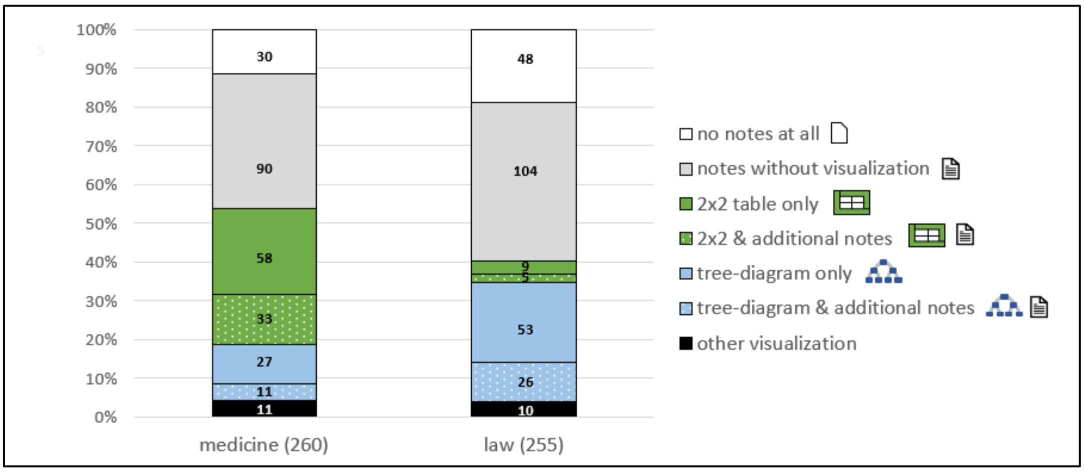


Figure 1: proportion of visualizations used in the notes of the medical and law students

Generally, only 2.5% of the participants used the highlighting principle. Figure 2 displays the further characteristics of the representations by illustrating the proportion of a) the numerical representations (coded among the frequency strategy) and b) the relevant complete and correct information in the representations (coded among the completion strategy in the questions 1-4 and 7).

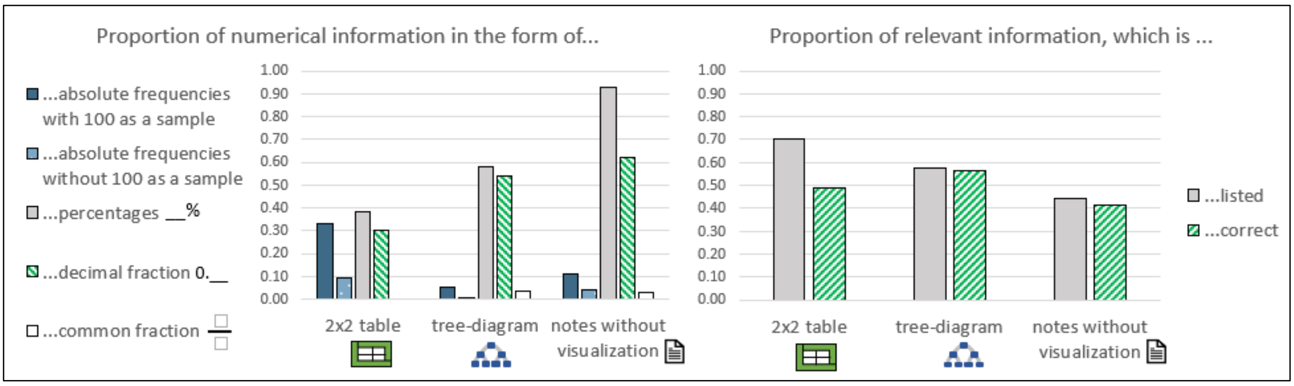


Figure 2: characteristics of the different representations in the student's notes

Discussion

The results show that students hardly ever use the frequency strategy for working on the calculation of a Bayesian situation without any instruction. The most common numerical representations are percentages and decimal fractions in all representations. This generally confirms the results by Weber et al. (2018). In the 2×2 table absolute frequencies are used more often than in other visualization, which could be based on the fact, that frequency 2×2 tables are used in German schools (contrary to frequency tree-diagrams). However, in the frequency 2×2 tables, the sample is often 100 which makes it ambiguous, if students actually wanted to use absolute frequencies, or mistakenly omitted the percentage sign "%". Thus, these results could show that it is not intuitive to use the more supportive natural frequencies and explicit training to do so is required. Moreover, the results show that 54% of the medical and 40% of the law students use the visualization strategy. However, the probability tree-diagram and 2×2 table are almost the only visualizations which are used. This might be illustrative of the current German curriculums in which these visualizations are regularly implemented, but not more promising ones such as a frequency double-tree, frequency net or unit square. Interestingly, there seems to be a difference in preference for the choice of visualization between the medical and law students with more medical students using the 2×2 table. This may be based on self-studying of the medical students with material with 2×2 tables for the medical state exam, (compare https://www.amboss.com/de/wissen/Epidemiologie_und_Wahrscheinlichkeiten/). Additionally, it stands out that using the 2×2 table is frequent, but adding the numerical information into the 2×2 table is error-prone. This is mainly based on the fact that the given conditional probabilities (sensitivity and false-positive rate especially) are added in the 2×2 table (as in Table 3), where joint probabilities would be required. This underlines arguments about 2×2 tables being misguiding, as given conditional probabilities cannot be represented (compare Büchter et al., 2022).

The results call for further analyses. For instance, additional analyses could be carried out to study the association between the strategies which were applied and the performance of calculation. Moreover, in further analyses it will be interesting to identify if and how the used strategies can be changed through participation in a training course on Bayesian reasoning and further if the initial use of the strategies can help to clarify which of the training methods is most supportive.

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*, 1186.

Binder, K., Krauss, S., & Wiesner, P. (2020). A New Visualization for Probabilistic Situations Containing Two Binary Events: The Frequency Net. *Frontiers in Psychology*, *11*, 750.

Binder, K., Steib, N., & Krauss, S. (2022). Von Baumdiagrammen über Doppelbäume zu Häufigkeitsnetzen – kognitive Überlastung oder didaktische Unterstützung? *Journal für Mathematik-Didaktik*, 1–33. https://doi.org/10.1007/s13138-022-00215-9

Böcherer-Linder, K., & Eichler, A. (2019). How to Improve Performance in Bayesian Inference Tasks: A Comparison of Five Visualizations. *Frontiers in Psychology*, *10*, 267.

Büchter, T., Eichler, A., Steib, N., Binder, K., Böcherer-Linder, K., Krauss, S., & Vogel, M. (2022). How to Train Novices in Bayesian Reasoning. *Mathematics*, *10*(9), 1558.

Clinton, V., Morsanyi, K., Alibali, M. W., & Nathan, M. (2016). Learning about Probability from Text and Tables: Do Color Coding and Labeling through an Interactive-user Interface Help? *Cognitive Psychology*, *30*(3), 440–453. https://doi.org/10.1002/acp.3223

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

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Communicating statistical information. *Science (New York, N.Y.)*, *290*(5500), 2261–2262. https://doi.org/10.1126/science.290.5500.2261

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.

Mayring, P. (2015). Qualitative Content Analysis: Theoretical Background and Procedures. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to Qualitative Research in Mathematics Education* (pp. 365-380). Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9181-6_13

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

Micallef, L., Dragicevic, P., & Fekete, J.-D. (2012). Assessing the Effect of Visualizations on Bayesian Reasoning through Crowdsourcing. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2536–2545. https://doi.org/10.1109/TVCG.2012.199

Weber, P., Binder, K., & Krauss, S. (2018). Why Can Only 24% Solve Bayesian Reasoning Problems in Natural Frequencies: Frequency Phobia in Spite of Probability Blindness. *Frontiers in Psychology*, *9*, 1833. https://doi.org/10.3389/fpsyg.2018.01833

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition*, *98*(3), 287–308.