HOW TO UNDERSTAND COVARIATION IN BAYESIAN REASONING SITUATIONS WITH DOUBLE-TREES AND UNIT SQUARES

Karin Binder[1], Markus Vogel[2], Katharina Böcherer-Linder[3], Theresa Büchter[4], Andreas Eichler[4], Stefan Krauss[5], and Nicole Steib[5]
[1] Ludwig Maximilian University Munich, Germany
[2] University of Education Heidelberg, Germany
[3] Albert-Ludwigs-University Freiburg, Germany
[4] University of Kassel, Germany
[5] University of Regensburg, Germany
Karin.Binder@lmu.de

There have been intensive research efforts to improve Bayesian reasoning over the last 25 years. Much of this research focuses solely on improving performance on Bayesian tasks. In addition to performance, however, it is also important to establish an understanding of the effect on the positive predictive value when parameters of Bayesian formula are changed. We call this ability "covariation" in Bayesian tasks. To this end, training courses were developed to support understanding of covariation, based on strategies that have been proven helpful by previous studies concerning performance, using: (a) natural frequencies and (b) visualisations, i.e., double trees and unit squares. Results of a comparative study in a pre-, post-, and follow-up test design show that the developed training courses can improve understanding of covariation.

INTRODUCTION

Nowadays, problem situations of Bayesian reasoning have become familiar (consciously or not) among the general public, thanks to the coronavirus pandemic. Imagine the following situation: You have just returned from a high incidence area with symptoms of a cold and used the "AESKU.RAPID" self-test to identify whether you have been infected with SARS-CoV-2. You have tested positive. The following statistics on individuals who have likewise just returned from a high incidence area with symptoms of a cold, and on the AESKU.RAPID self-test, reveal the following.

- There is a 5% probability that a person is infected with SARS-CoV-2 (base rate).
- If a person is infected with SARS-CoV-2, then the probability is 96% that the person tests positive (sensitivity).
- If a person is *not* infected with SARS-CoV-2, then the probability is 2% that the person tests positive nevertheless (false-positive rate).

Questions such as, "If a person tests positive, then what is the probability that the person is infected with SARS-Cov-2 (positive predictive value)?" can be answered using Bayes formula. Corresponding tasks are therefore called "Bayesian tasks." Here, the solution for the Bayesian task could be calculated by $P(\text{infected}|\text{test positive}) = \frac{5\%\cdot 96\%}{5\%\cdot 96\% + 95\%\cdot 2\%} = 71.6\%$. However, numerous studies have shown that even experts (from different professions) fail to calculate it this way and fail to answer comparable questions (Eddy, 1982; Sloman et al., 2003). Yet almost all corresponding empirical research from cognitive psychology and mathematics education exclusively focuses on the *performance* of participants in such situations, thus ignoring the fact that parameters might change.

As part of the TrainBayes project (http://www.bayesianreasoning.de/en/br_trainbayes_en.html), we developed different training courses and examined *three* different aspects of Bayesian reasoning:

1. *Performance:* The ability to calculate the positive predictive value, if e.g., base rate, sensitivity, and false-positive rate are provided.
2. *Covariation:* The ability to appropriately estimate the effect of changing parameters (such as base rate, sensitivity, or false-positive rate) on the positive predictive value.
3. *Communication:* The ability to adequately communicate the results from Bayes formula in an expert–layman setting.

In this paper, we focus on the effect of training courses to improve the Bayesian reasoning skills of medical and law students. We primarily focus on development of the "covariation" aspect. The implemented training methods can be found in Büchter, Eichler, et al. (2022), and considerations for the

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org ©2022 ISI/IASE

development of appropriate visualisations with respect to multimedia aspects can be found in Büchter, Steib, et al. (2022).

BACKGROUND

Much existing research on Bayesian reasoning focuses mainly on performance. Even though many studies show that people fail to successfully answer Bayesian tasks, Bayesian reasoning can be improved by using certain strategies. The meta-analysis by McDowell and Jacobs (2017) shows that by using (a) *so-called "natural frequencies"* (i.e., a pair of two absolute frequencies such as "5 out of 100 people") and (b) *visualisations*, more people answer the above question correctly (see also Gigerenzer & Hoffrage, 1995). The task could be translated into natural frequencies as follows.

- Fifty out of 1,000 individuals are infected with SARS-CoV2 (base rate).
- Out of the 50 individuals who are infected with SARS-CoV-2, 48 test positive (sensitivity).
- Out of the 950 individuals who are *not* infected with SARS-CoV-2, 19 test positive (false-positive rate).

In the format of natural frequencies, the answer to a question such as "How many of the individuals who test positive with a self-test are actually infected with SARS-Cov-2 (positive predictive value)?" is much easier to see. The solution algorithm simplifies (Gigerenzer & Hoffrage, 1995) and is given here by the ratio $\frac{48}{48+19}$, which yields the answer 48 out of 67. Furthermore, appropriate visualisations can help people generate the solution to the task (for an overview, see, e.g., Binder et al., 2015; Binder et al., 2021; Eichler et al., 2020; Khan et al., 2015; Spiegelhalter et al., 2011). In the field of Bayesian reasoning, many visualisations are used (e.g., double-tree diagrams, unit square, 2 × 2 tables, tree diagrams, icon arrays). Some of these visualization support solution-finding better than others.

Two visualisations that have proven empirically fruitful for teaching Bayesian reasoning in the past are double trees (Binder et al., 2020; Martignon & Kuntze, 2015; Wassner et al., 2004) and unit squares (Böcherer-Linder & Eichler, 2017; Pfannkuch & Budgett, 2017). Figure 1 shows a double tree (left) and a unit square (right) for a general medical problem, which is structurally identical to the aforementioned SARS-CoV-2-problem. We used these visualizations and this general medical problem in a training course to support developing understanding of covariation, as described below.

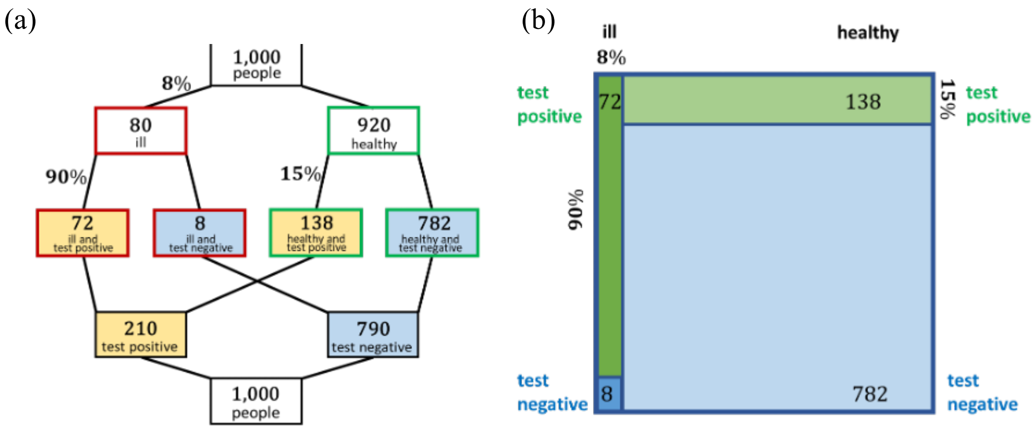


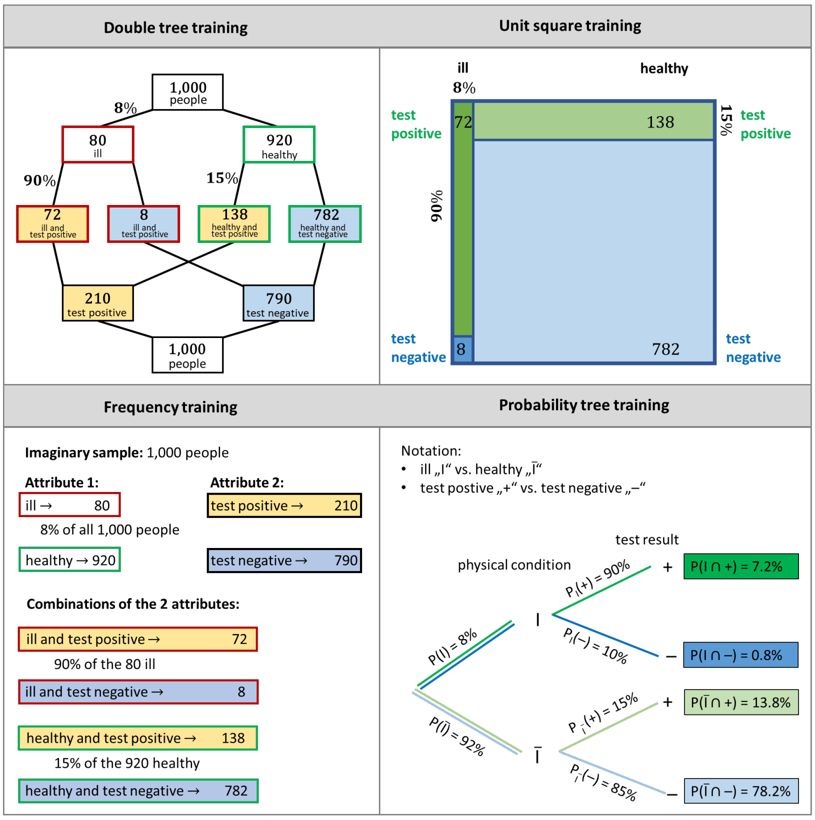
Figure 1. (a) Double tree and (b) unit square for a general medical problem with a base rate of 8%, sensitivity of 90%, and false-positive rate of 15% (This universal context was used in the training course.)

To reach a comprehensive understanding of Bayesian situations, more is required than just being able to calculate a correct solution using Bayes' formula: an understanding of covariation is also needed. Also, Borovcnik (2012) demands that opportunities should be created to "investigate the influence of variations of input parameters on the results" (p. 21). To our best knowledge, however, very few research studies such as Böcherer-Linder, Eichler, and Vogel (2017) have addressed this issue so far.

DESIGN OF THE STUDY

As part of the TrainBayes project, we developed digital learning environments to improve participants' ability to estimate the effect of changed formula parameters (covariation). An overview of previously developed training courses (with a focus on performance) can be found in Büchter, Eichler, et al. (2022). The main focus here was on two training courses, both based on the strategy of natural frequencies. One used double trees and the other used unit squares. In addition, in this study, a waiting control group and two different control training courses (see also Table 1) were implemented. The "frequencies" control training group only learned how to use natural frequencies in solving Bayesian problems but did not work with any visualisation aspects. The "probability tree" control training group worked with tree diagrams containing probabilities but did not use the natural frequencies strategy.

Table 1. Overview of the four different training courses. Furthermore, there was another control group without any training involved.



TRAINING COURSES FOR IMPROVING COVARIATION

Covariation requires basic knowledge in solving a Bayesian problem as described above. Therefore, the aspect of *performance* was trained first. The covariation training included explanatory videos as well as dynamic visualisations with sliders that enable independent discovery of the influences (see Figure 2). Based on the visualisation of the Bayesian situation in the double tree, users can observe how the absolute frequencies in the nodes of the double tree change when individual parameters change. Users also can observe how changes in the absolute frequencies affect the numerator and denominator of the fraction, and thus the positive predictive value. In an exercise phase, students worked on the learning task on SARS-CoV-2 self-tests introduced at the beginning of this paper. To create an authentic scenario where the given parameters of this Bayesian situation might change, the students were, for instance, asked to imagine using a different SARS-CoV-2 self-test with a smaller false-positive rate (compared to the false-positive rate of the AEKSU.RAPID test). As a consequence, they assessed the impact of changes to the parameter (false-positive rate, sensitivity, or base rate) on the positive predictive value, and also on the two conjunctive probabilities that are required to calculate this value.

An analogous training course was developed for the covariation training using unit squares. Again, corresponding sliders are located within the dynamic visualisation to make the effects of parameter changes directly perceivable.

Sensitivity increases	
1. How do the relevant frequencies change?	**2. How does the fraction change?**
• Frequency of the ill people who test positive increases • Frequency of the healthy people who test positive stays the same	• Numerator: increases • First summand of the denominator: increases • Second summand of the denominator: stays the same • Altogether: numerator and denominator increase by the same value
Conclusion: Fraction (thus the positive predictive value) increases	

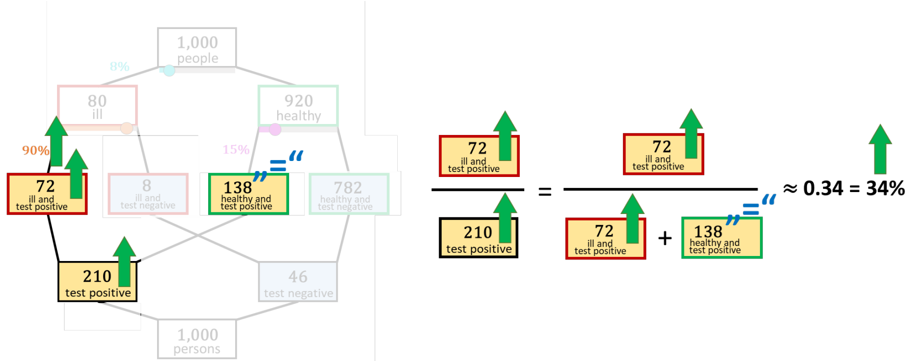


Figure 2. Excerpt from the double tree training to improve the aspect of covariation

METHOD

The effectiveness of training courses within our digital learning environments was examined in a pre-, post-, and follow-up test design. The follow-up test was conducted 8–10 weeks after the post-test. The training course took place between pre- and post-test and lasted approximately 30 minutes. Our participants were 260 medical and 255 law students who worked on contexts related to their chosen professions (i.e., medical students on medical contexts and law students on legal contexts). Participation in the study was voluntary; written informed consent was obtained from the participants, and they received payment for participation.

In the pre-, post-, and follow-up tests, participants' covariation was measured in the following way: subjects were told that, for example, in the aforementioned situation (SARS-CoV-2) there is a self-test with a *lower false positive rate*. The subjects then had to decide how this lower false positive rate affects:

- the probability of being infected AND receiving a positive test result;
- the probability of *not* being infected AND receiving a positive test result; and
- the probability of being infected if the test result is positive.

We asked about the effect of the change on both relevant joint probabilities and on the positive predictive value. Answer options were: becomes smaller, remains the same, and becomes larger. If the correct answer was given for the change of the positive predictive value, the participant was also presented with a reason for this change, which had to be evaluated as correct or incorrect. The participant was given a second one if this decision was correct again. This type of testing for covariation was done not only in terms of a false-positive rate, but also in terms of sensitivity and base rate. An overview of the two training courses with double trees and unit squares can be found in the form of screenshots from the explanatory videos in supplementary material that can be downloaded at https://osf.io/tkh5s/.

RESULTS

Test surveys were recently completed, and data analysis is still ongoing. Therefore, we cannot present final quantitative results, but initial insights into the data are offered below.

From analysing the results of how participants judged changes to the positive predictive value (third single-choice question), it can be noted that the solution rates were already remarkably high in the pre-test for changes in the false-positive rate (65.8%) and for changes in the sensitivity (69.3%) across all training groups—even considering a 33% probability of guessing the correct answer. Furthermore, the solution rates for changes in both the false-positive rate and sensitivity increase in the post-test and follow-up test in comparison with the pre-test, but without notable differences between training and waiting control groups (see Table 2). However, the training courses substantially improve judgment for changes in the base rate (see Table 2). For changes in the base rate, the solution rate in the pre-test is lower than for changes in the false-positive rate or sensitivity. Yet, participants in the training courses showed substantially higher solution rates, with a slight drop in the follow-up test.

Table 2. Solution rates for judging changes in the base rate, false-positive rate, or sensitivity

	Pre-test	Post-test	Follow-up test
Base rate			
Waiting control group (103 participants)	40.8%	52.4%	48.5%
Training groups (412 participants)	41.7%	68.9%	61.9%
False-positive rate			
Waiting control group (103 participants)	66.0%	78.6%	75.7%
Training groups (412 participants)	65.8%	78.9%	74.8%
Sensitivity			
Waiting control group (103 participants)	70.9%	73.8%	79.6%
Training groups (412 participants)	68.0%	79.4%	76.9%

Figure 3 shows how students approach such covariation tasks after training and proceed step by step to the solution. They consider how parameter changes affect the absolute frequencies in the double tree and check how these changes affect the numerator and denominator of the positive predictive value.

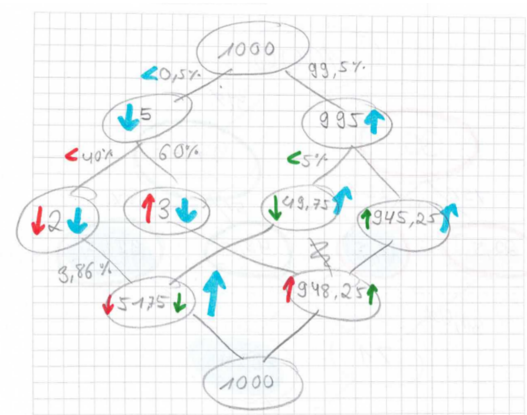


Figure 3. Task processing by a participant after the covariation training

CONCLUSION

Although Bayesian reasoning has been studied mainly in relation to the aspect of "performance," we wanted to broaden this field and take a closer look at the aspect of covariation. The present study gives initial insights into how training on the aspect of covariation can be provided and that appropriate training can improve this aspect—particularly for the influence of the base rate. Research on covariation should be pursued further because it sheds light on a previously neglected aspect of Bayesian reasoning, which is an essential competence in dealing with statistical information. Furthermore, such trainings should find their way into learning scenarios in various university courses.

FUNDING

The research was funded by the DFG—German Research Foundation (EIC773/1-4).

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,* Article 1186. https://doi.org/10.3389/fpsyg.2015.01186

Binder, K., Krauss, S., Schmidmaier, R., & Braun, L. T. (2021). Natural frequency trees improve diagnostic efficiency in Bayesian reasoning. *Advances in Health Sciences Education, 26*(3), 847–863. https://doi.org/10.1007/s10459-020-10025-8

Binder, K., Krauss, S., & Wiesner, P. (2020). A new visualization for probabilistic situations containing two binary events: The frequency net. *Frontiers in Psychology, 11*, Article 750. https://doi.org/10.3389/fpsyg.2020.00750

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*, Article 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), 25–46. https://doi.org/10.35763/aiem.v1i11.169

Borovcnik, M. (2012). Multiple perspectives on the concept of conditional probability. *Avances de Investigación en Educación Matemática,* (2), 5–27. https://doi.org/10.35763/aiem.v1i2.32

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), Article 1558. https://doi.org/10.3390/math10091558

Büchter, T., Steib, N., Böcherer-Linder, K., Eichler, A., Krauss, S., Binder, K., & Vogel, M. (2022). *Designing visualizations for Bayesian problems according to Multimedia Principles* [Manuscript submitted for publication]. University of Kassel.

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 University Press.

Eichler, A., Böcherer-Linder, K., & Vogel, M. (2020). Different visualizations cause different strategies when dealing with Bayesian situations. *Frontiers in Psychology,* (11), Article 1897. https://doi.org/10.3389/fpsyg.2020.01897

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

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

Martignon, L., & Kuntze, S. (2015). Good models and good representations are a support for learners' risk assessment. *The Mathematics Enthusiast, 12*(1), 157–167. https://doi.org/10.54870/1551-3440.1341

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

Pfannkuch, M., & Budgett, S. (2017). Reasoning from an eikosogram: An exploratory study. *International Journal of Research in Undergraduate Mathematics Education*, *3*, 283–310. https://doi.org/10.1007/s40753-016-0043-0

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, *91*(2), 296–309. https://doi.org/10.1016/S0749-5978(03)00021-9

Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science, 333*(6048), 1393–1400. https://doi.org/10.1126/science.1191181

Wassner, C., Martignon, L., & Biehler, R. (2004). Bayesianisches denken in der schule. *Unterrichtswissenschaft, 32*(1), 58–96. https://doi.org/10.25656/01:5808