From research on Bayesian reasoning to classroom intervention

Desde la investigación sobre razonamiento Bayesiano a la intervención en el aula

Andreas Eichler[1], Katharina Böcherer-Linder[2] and Markus Vogel[3]

[1]University of Kassel, [2]University of Freiburg,

[3]University of Education Heidelberg, Germany

Abstract

Dealing with Bayes' rule is the mathematical part of judgement in situations of uncertainty. These situations are of importance for crucial judgements in medicine, law and further professions. Since laymen and experts have severe difficulties of applying Bayes' rule, the question how to facilitate dealing with Bayesian situations, i.e. situations in which Bayes' rule could be applied is posed. Our research built upon the well-established facilitating strategy of using natural frequencies as information format in Bayesian situations. On this basis, we have investigated different visualizations and developed a training of dealing and understanding Bayesian situations. Our results suggest that the unit square and the double tree diagram are appropriate visualizations for a training concerning Bayesian situations and that also a brief training has strong effects.

Keywords: Bayes' rule, conditional probability, visualization, training, learning

Resumen

La parte matemática del razonamiento en situaciones de incertidumbre implica el uso del teorema de Bayes. Estas situaciones son importantes para la emisión de juicios en medicina, derecho y otras profesiones. Puesto que tanto las personas ordinarias como los expertos tiene dificultades severas para aplicar el teorema de Bayes, se plantea la cuestión de cómo facilitar el tratamiento de las situaciones de Bayes, esto es, situaciones en las que la regla de Bayes puede ser aplicada. Nuestra investigación se basa en la estrategia facilitadora bien establecida de usar frecuencias naturales como formato de información en situaciones Bayesianas. Sobre esta base, hemos investigado diferentes visualizaciones y desarrollado una intervención formativa para tratar y comprender situaciones Bayesianas. Nuestros resultados sugieren que el cuadrado unitario y el doble diagrama en árbol son visualizaciones apropiadas para el entrenamiento relativo a las situaciones Bayesianas y que incluso un breve entrenamiento tiene fuertes efectos.

Palabras clave: Teorema de Bayes, probabilidad condicional, visualización, enseñanza, aprendizaje

1. Introduction

Conditional probabilities entail severe difficulties (Diaz, Batanero, & Contreras, 2010). Research in psychology and mathematics education refer to different issues of conditional probabilities (Böcherer-Linder, Eichler, & Vogel, 2018a; Diaz & Batanero, 2009; Diaz et al., 2010; Kahneman, Slovic, & Tversky, 1982), but often refer to the specific topic of Bayes' rule that represents an obstacle for students or laymen as well as experts (Gigerenzer & Hoffrage, 1995). However, Bayes' rule is important for dealing with judgement and decision making in situations of uncertainty, e.g. Bayesian situations. Johnson and Tubau (2015) provide a typical Bayesian situation in an unspecific medical context as follows:

Eichler, A. Böcherer-Linder, K. y Vogel, M. (2019). From research on Bayesian reasoning to classroom intervention. En J. M. Contreras, M. M. Gea, M. M. López-Martín y E. Molina-Portillo (Eds.), *Actas del Tercer Congreso Internacional Virtual de Educación Estadística.* Disponible en www.ugr.es/local/fqm126/civeest.html

> 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?

This Bayesian situation requires a judgement in an unspecific medical context. For an adequate judgement the nested sets of *disease and positive* and *positive* must be identified and processed as $P(disease|positive) = \frac{10\% \cdot 60\%}{10\% \cdot 60\% + 90\% \cdot 20\%}$. Several realistic medical contexts illustrate the importance of Bayes' rule as mathematical for judgement in uncertain situations (Steckelberg, Balgenorth, Berger, & Mühlhauser, 2004). Also in further professions, Bayes' rule is the basis for crucial judgements (Hoffrage, Hafenbrädl, & Bouquet, 2015). For example a similar situation is represented by the judgement of a lawyer about being guilt or innocence (Satake & Murray, 2014). As a consequence of the importance of Bayes' rule for making adequate judgements in real life situations, it is worthwhile to investigate the mentioned obstacles of people and to develop strategies to improve people's reasoning within Bayesian situations.

Up to now, research yield two strategies that facilitate dealing with Bayesian situations, i.e. to use natural frequencies and to use visualizations (McDowell & Jacobs, 2017). The first strategy is well documented. Following the results in the meta-analysis of McDowell and Jacobs (2017), natural frequencies increase the performance of people in Bayesian situations from about 5% to about 25%. Referring the situation given above, the statistical information in form of natural frequencies represents the sampling process from a virtual number of people as follows:

> 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 proportion of having the disease given that the test is positive?

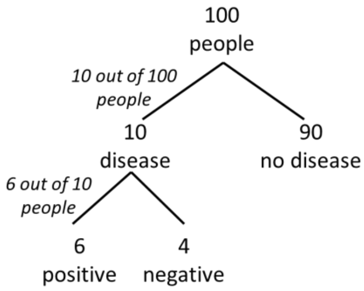
Also in this situation, the woman who have the disease and are positive tested and the women who are positive tested must be identified to compute $P(disease|positive) = \frac{6}{6+18}$.

Further, the strategy of visualizing the statistical information in a Bayesian situation facilitates dealing with these situations. However, although studies using visualizations in addition to natural frequencies reported an increase peoples performance in Bayesian situations to about 40% to 70% (Binder, Krauss, & Bruckmaier, 2015; Böcherer-Linder & Eichler, 2017), the facilitating effect of visualization is not as clear as the facilitating effect of natural frequencies (Johnson & Tubau, 2015; McDowell & Jacobs, 2017).

A third strategy, i.e. a training based on the first two strategies, was seldom focused in research: "Recent research has […] overlooked the underpinning mechanism of the training effects" (Sirota, Kostovičová, & Vallée-Tourangeau, 2014, p. 7). For this reason, a main goal in our research program BAYES[2] is to develop a training concerning Bayesian reasoning that could be used with students or laymen as well as with learners for a specific profession like medicine or law. A first decision for this training was to start with natural frequencies as information format. Since the results referring an optimal visualization are ambiguous, we started research in the field of visualizing Bayesian situations that we describe in the next section. Afterwards we describe the development of a learning environment for improving Bayesian reasoning and discuss results of different piloting studies with this learning environment.

2. Visualizing Bayesian situations

The most common visualizations of Bayesian situations in school are tree diagrams and 2x2-tables (e.g. Veaux, Velleman, & Bock, 2012). However, the 2x2-table is not a visualization of a Bayesian situation as given above, but only a well-structured representation of the results after dealing with a Bayesian situation. For example, given the base rate ("10 out of 100 women at age forty who participate in a study have a particular disease") and the first conditional probability ("60% of women with the disease will have a positive reaction to a test") could be represented in a tree (Fig. 1). However, it could not be represented in a 2x2-table without further considerations since a 2x2-table only contains the natural frequency or the probability of a compound event but not a conditional probability.



	disease	no disease	sum
positive	6		
negative			
sum	10	90	100

Figure 1. Visualization of conditional probabilities in the tree diagram and the 2x2-table

However, a further development of a 2x2-table, i.e. the unit square (Fig. 2) that provides an area proportional visualization of Bayesian situations represents both conditional probabilities as length of segments and probabilities of compound events as areas within the square or, respectively, as result of $P(disease) \cdot P(positive|disease)$.

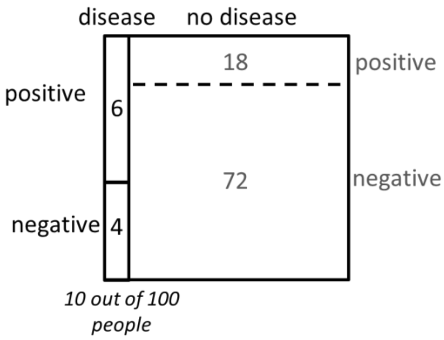


Figure 2. Visualization of conditional probabilities in the unit square

The unit square was proposed for visualizing conditional probabilities or Bayesian situations in educational contexts in different countries with different names (eikosogram: Budgett, Pfannkuch, & Franklin, 2016; Oldford, 2003; unit square: Eichler & Vogel, 2013; mosaic display: Friendly, 1999).

2.1. Investigation of the effectiveness of the tree diagram and the unit square

According to the considerations above, we investigated the effectiveness of two visualizations of Bayesian situations in an educational context (Böcherer-Linder et al., 2018a). For this we firstly analyzed people's performance when dealing with Bayesian situations supported by one of the two visualizations (Böcherer-Linder & Eichler, 2017). In this research we divided a group 143 undergraduate students randomly in two conditions, a tree-group and a unit square-group. Each group got a brief instruction (one page) how to use the tree diagram or, respectively, the unit square. Afterwards the students were asked to solve four problems referring to four different Bayesian situations. One of the items is given in Figure 3.

Medical diagnosis test
In a preventive medical check-up, 1000 people are tested. The test has the following characteristic: 80% of the infected people and 10% of the uninfected people get a positive test result.
Calculate the proportion of infected people among those testing positive. Write as a fraction.

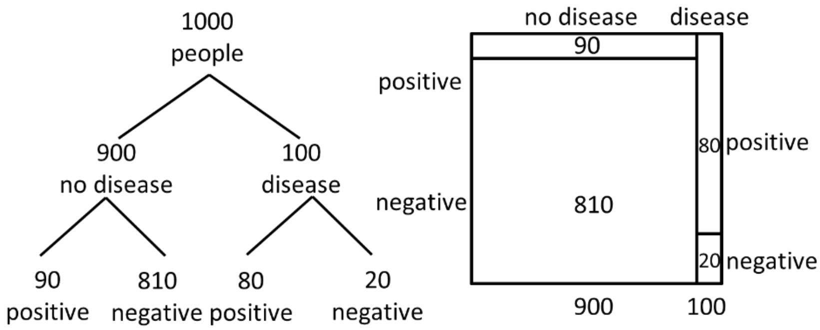


Figure 3. Sample item. Each group got only one visualization, i.e. the tree diagram or the unit square

Since the results showed a high reliability referring Cronbach’s alpha (α > 0.8), we added the results for each item that was 1 (correct solution) or 0 (false solution). Concerning this accumulated score, we found a significant supremacy of unit square (M= 2.93, D=1.417) compared to the tree diagram (M=1.72, SD=1.494), t(141)=4.961, $p < 0.001$) with a large effect size (Cohen’s d=0.84).

In this research, we found strong evidence that the degree of making the nested-set structure of a Bayesian situation transparent (Sloman, Over, Slovak, & Stibel, 2003) is crucial for the effectiveness of a visualization. Actually the unit square makes this structure transparent since the relevant sets (disease and positive; no disease and positive) that must be identified to apply Bayes’ rule are represented by neighbored areas. By contrast, these areas are represented by different paths in the tree diagram and does not represent the hierarchical structure of the tree diagram (cf. Böcherer-Linder & Eichler, 2017; Fig. 4).

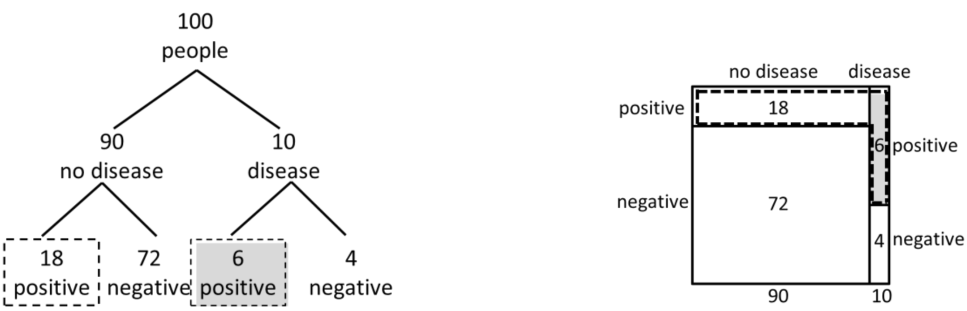
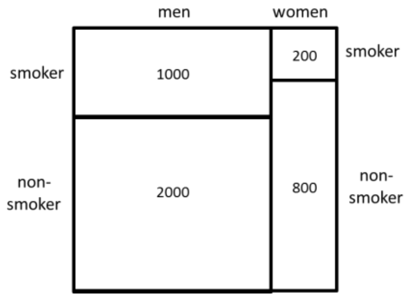


Figure 4. Visualization of conditional probabilities in the tree diagram and the 2x2-table

In an additional study (Böcherer-Linder, Eichler, & Vogel, 2017), we addressed the view that solving a problem in a Bayesian situation like given above using Bayes formula is only a part of understanding a Bayesian situation (Borovcnik, 2012). For this reason, we developed items to investigate whether people are able to estimate the effect of changing parameters like the base rate in Bayesian situations. Since a parameter change results in a change of different other probabilities in a Bayesian situation, we called this aspect of understanding a Bayesian situation covariation aspect. One item for the covariation aspect is given in Figure 5.

Smoke
4000 students of a university were asked if they smoke or not. It turned out that one-third of the men smoke and one-fifth of the women smoke:



How the following proportions change if, one year later, there are more women among the 4000 students of the university and the smoking behavior of men and women is still the same?

Mark the correct solutions.
The percentage of non-smokers among the women will be bigger / smaller / constant.
The percentage of women among the smokers will be bigger / smaller / constant.
The percentage of men among the non-smokers will be bigger / smaller / constant.

Figure 5. Sample item for the covariation aspect

In two studies with undergraduate students (n=148 and n=143), we divided the students in two groups as explained above. We used three different contexts and found again a significant supremacy of the unit square although the effect size was small (Cohen's d=0,34). However this result could be explained by the items itself since the probability for having the correct solution by guessing is 1/3.

2.2. Investigation of the effectiveness of five visualizations

In psychological research, further visualizations of Bayesian situations are investigated (e.g. Binder, Krauss, & Bruckmaier, 2015). For this reason, we enhanced our focus on visualizing Bayesian situations according to three properties of visualizations that were found to be effective (Böcherer-Linder & Eichler, in review):

- As mentioned above, the degree of making the nested-sets structure of a Bayesian situation transparent was found to be an effective property of a visualization (e.g. Böcherer-Linder & Eichler, 2017; Sloman et al., 2003)
- Different researchers also investigated the facilitating effect of using representations of "real, discrete and countable" objects (Cosmides & Tooby, 1996, p. 33) and found a significant effect (e.g. Brase, 2009; Garcia-Retamero & Hoffrage, 2013).
- Finally, an area-proportionality was found as facilitating factor of visualizing Bayesian situations (e.g. Talboy & Schneider, 2017; Tsai, Miller, & Kirlik, 2011).

We built upon our former research and proved hypotheses concerning the effect of these three properties on the basis of the tree diagram and the unit square as shown in Table 1.

Table 1. Three hypotheses concerning three properties of visualizations

Property	Given in	Not given in	Hypothesis
Area proportionality	Unit square	2x2-table	The unit square is more effective than the 2x2-table
Discrete objects	Icon array	Unit square	The icon array is more effective than the unit square
Nested-sets transparency	Double tree	Tree diagram	The tree diagram is more effective than the double tree

The five visualizations indicated in Table 1 are shown in Fig. 6 referring the medical diagnosis situation explained above.

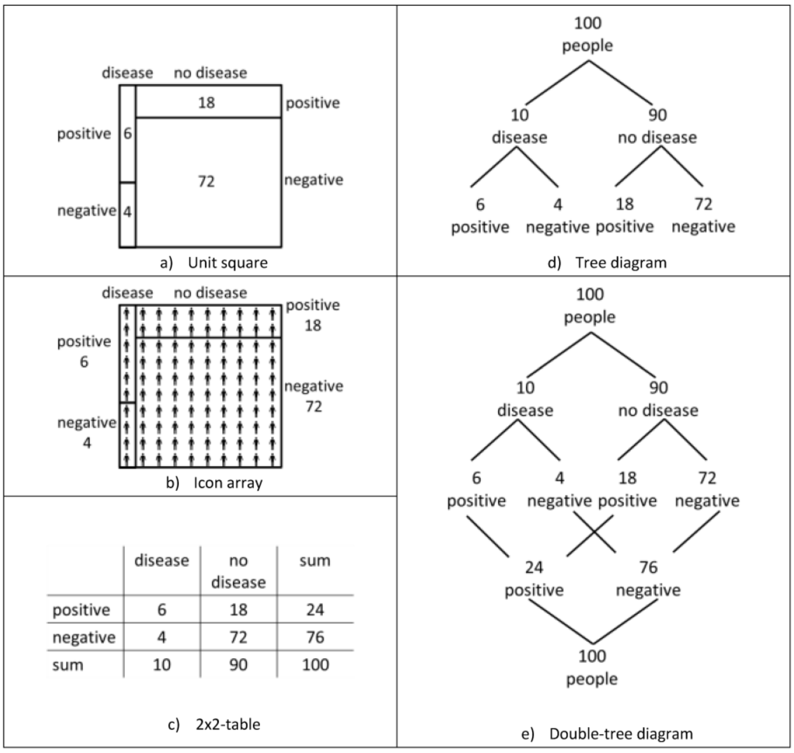


Figure 6. Five visualisations in a medical diagnosis situation

In an experiment with 688 undergraduate students, we used again four items that were identical to the study explained above (Böcherer-Linder & Eichler, 2017). As a result, we found a significant effect of the discrete objects since the icon array (M=2.56, SD=1.31) outperformed the unit square (M=2.26, SD=1.41; $t(294.238)=1.882$, $p<0.05$) with a small effect (Cohen's d=0.22). Further and unexpected the 2x2-table (M=2.76, SD=1.33) outperformed the unit square (M=2.26, SD=1.41; $t(293.619)=3.142$, two-tailed: $p< .01$) also with a small effect (Cohen's d=0.37). Finally, the double tree (M=2.03, SD=1.58) outperformed the tree diagram (32.2%; M=1.28, SD=1.34, $t(232.696) = 3.989$, $p < .001$) with a middle effect (Cohen's d=0.51).

In addition, post hoc analyses showed that any visualization is more effective than the tree diagram with mostly a large effect (comparison with the unit square, the icon array and the 2x2-table. The 2x2-table, the unit square and the icon array make the nested-sets structure in a Bayesian transparent as the double tree do. For this reason, our results give strong evidence that particularly making the nested-sets structure of a Bayesian situation transparent represents an effective property of visualizing Bayesian situations.

In the same research, we also investigated the covariation aspect with modified items. However, the data were not analyzed yet. For this reason we could not back our hypothesis that the property of an area proportionality facilitates to solve problems referring to the covariation aspect.

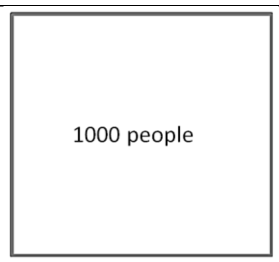
3. A training for improving Bayesian reasoning

Our research informed the following decision for developing an optimal training for Bayesian reasoning:

- The 2x2-table and the icon array outperformed the unit square. However, in a narrow sense, a 2x2-table is not a visualization of a Bayesian situation since it does not visualize the conditional probabilities in these situations (see above). The icon array is appropriate for representing Bayesian situations if the visualization is given. If the visualization has to be developed by a learner, the icon array seems to be not appropriate since in most cases the learner would be required to draw a huge amount of icons. For this reason, we defined the unit square as an appropriate visualization for training Bayesian reasoning.
- The double tree outperformed the tree diagram. Since we did not found differences in the performances of those students who used the double tree compared to the students that used the unit square, we also defined the double tree as an appropriate visualization for training Bayesian reasoning.

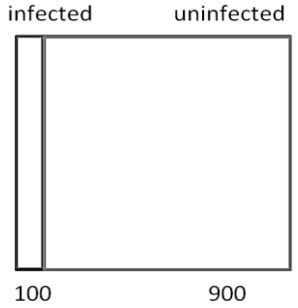
Following the idea of Sedlmeier and Gigerenzer (2001), we developed a training with an as much as possible little time for working with Bayesian situations. This training has the following three phases (Böcherer-Linder, Eichler, & Vogel, 2018b). The first phase lasting about 10 minutes include the solution of a problem with an worked example (Renkl, 2002).

<table>
<tr><td colspan="2">Problem presentation:
After travelling to a far country, you learn that an average of 10% of the travelers contracted a new kind of disease during their trip. The disease proceeds initially without any clear symptoms, therefore you don’t know whether you had been infected or not.
You learn that a medical test was developed which has the following characteristics:
80% of infected people get a positive test result (sensitivity of the test).
15% of not infected people get a positive test result (specificity of the test).
Finally, you decide to carry out the test and get a positive test result. What is the probability that you actually have the disease?</td></tr>
<tr><td colspan="2">Step 1: Choice of the sample size</td></tr>
<tr><td>For solving the problem, we first consider the question what the probabilities mentioned in the text imply for a concrete group of travelling people. We choose a sample size, for example 1000 people.

We represent the group of 1000 people by drawing a unit square.</td><td>
</td></tr>
</table>

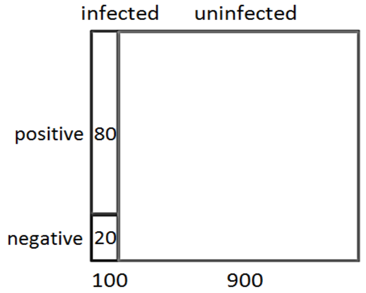
Step 2: Construction of the frequency representation

Since 10% of the travelling people contracted the disease during their trip, 100 out of the 1000 people are expected to be infected. 900 out of the 1000 people are expected to be uninfected.
Thus, we divide the unit square in vertical direction for “infected” and “uninfected” at the ratio of 100 to 900. At the bottom of the narrow rectangle we write “100” and at the bottom of the broader rectangle we write “900”.



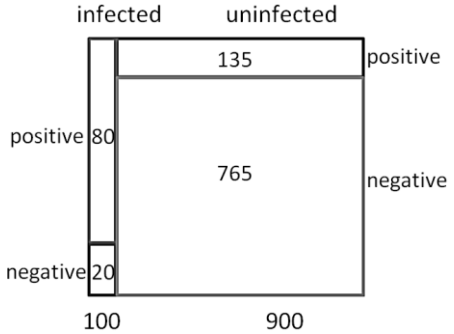
Since 80% of the infected people get a positive test result, 80 out of the 100 infected people are positively tested. Accordingly, 20 out of the 100 infected people are negatively tested.

Therefore we subdivide the narrow rectangle horizontally into two parts and write “80” and “20” into the resulting areas.



Since 15% of the uninfected people get a positive test result, 135 out of 900 uninfected people are positively tested (Because 15% of 900 is 135). Accordingly, the other 765 uninfected people get a negative test result.

Therefore we subdivide the broader rectangle horizontally into two parts for “positive” and “negative” and write the numbers into the resulting areas.

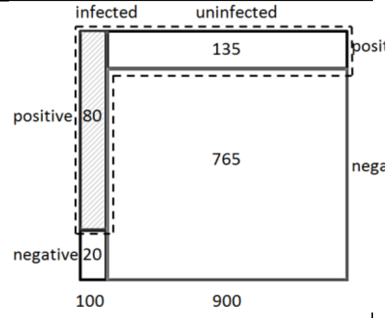


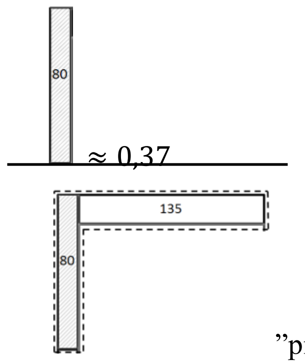
Requested is the probability that a person with a positive test result is actually infected.
Thus, we have to calculate which proportion of the positively tested people actually is infected.
For this aim, we surround all positively tested people with a dashed line in the unit square and emphasize with grey color all of them, that are infected.
We read out the following numbers:
Number of infected and positively tested: 80
Number of all positively tested: 80+135=215

We calculate the following proportion:

$$\frac{infected\ and\ positive}{all\ positively\ tested} = \frac{80}{80+135} \approx 0{,}37$$

This proportion corresponds to the requested probability of 37%.





”picture-formula”

Figure 7. Worked example in the training, phase 1

The second phase also lasting 10 minutes includes an exercise that was structurally identical to the problem in the first phase. After working with this problem with paper and pencil, the learners got a presentation of the correct solution. In this solution, we used a rough drawing of the unit square showing that the visualization is a tool for problem solving which did not necessarily imply a precise drawing.

In the third phase lasting 10 minutes, we addressed the covariation aspect and explain which probabilities are influenced if a parameter in a Bayesian situation changes. In figure 8 we illustrate the explanation referring to the change of the base rate.

3. Change of the base rate

The base rate indicates the proportion of having trisomy (5% of the children of women's of age 45). An increase of the base rate results in an increase of the area of the thin rectangular on the upper left side of the unit square and, accordingly, a decrease of the area of the rectangular on the upper right side.

If the base rate increases to 10%, there would be 90 true-positive people and 90 false-positive people. The change of the base rate changes the proportion as shown below:

$$\frac{90}{90+90} = \frac{90}{180} \approx 50\ \%$$

Actually, the result shows a considerable change of the probability.

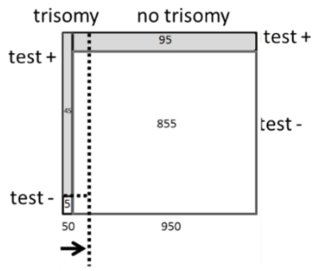


Figure 8. Part of the worked example in the training, phase 3.

4. Trainings

We conducted different trainings with different small groups and, finally, different research questions. In these pilot studies we restricted our focus currently on the unit square.

In a first pilot study, we tested our training with a group of 38 students in two courses of grade 11. One treatment group including 22 students got the training as shown above. The other 16 students form a control group. The design of the quasi-experiment is shown in Fig. 9. Since there were not all of the students present in the three phases shown in Fig. 5, our analysis is based on 16 students (treatment group) and 13 students (control group).

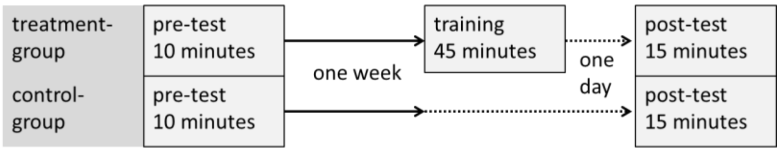


Figure 9. Design of the first pilot study

The pre-test and the post-test consisted of two Bayesian situations. One of the situations in both tests was the same, the other situation differed to investigate whether repeating a task has an effect on performance in Bayesian situations. In the pre-test, one situation included only a performance-task and the second situation included both a performance task and a task representing the covariation aspect. This task is as follows:

There are tests for diagnosing people if they have infectious diseases like measles or scarlet. Concerning such an infectious disease and the corresponding test the following information is given:

The probability of having the infectious disease is 2%. Given there is a patient having the disease, the test yields in 90% of all cases a “positive” result, which means it indicates correctly the infectious disease. Given there is a non-infected patient, the test shows in 5% of all cases also a “positive” result, which means it indicates the infectious disease by mistake.

a) What is the probability of a patient having actually the disease given a “positive” test result?

b) How changes the probability of section a) if the probability of having the disease is higher?

The main question in this pilot study was whether it is possible to train Bayesian reasoning effectively in a very short time.

In the second pilot study the sample consist of 19 master students that were randomly divided into two treatment groups. The first group got the training as shown above. The second group got the same training in which every natural frequency was substituted by a probability. The test was exactly the same as in the first pilot study.

5. Results

The reliability of the few test items in the pre-test and the post-test were appropriate (pre-test: Cronbach’s α = 0.63; post-test: Cronbach’s α = 0.85). For this reason, we investigated the accumulated test score (Min = 0, Max = 2).

Although the numbers of participant were small, we applied a mixed ANOVA (within-factor: performance in the pre-test and post-test; between factor: group) as a heuristic to investigate the training effect in the treatment group compared to the control group.

Figure 10 contain the results in this study referring the students’ performance in the Bayesian situations (solution of the item a) that show a clear effect of the treatment referring the means.

	Mean	Std.Dev.	N
Control Pre	.08	.13	13
Treatment Pre	0.19	.16	16
Control Post	.38	.13	13
Treatment Post	1.75	.16	16

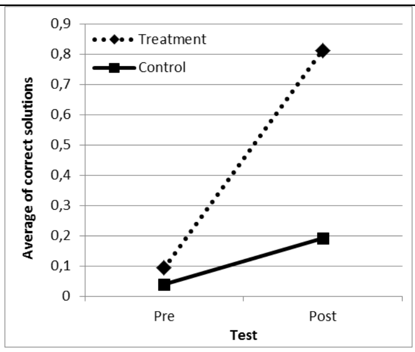


Figure 10. Descriptive results of the training concerning item a)

As expected, the ANOVA yields a significant effect (F=20,733, p=0,000) of the treatment with a strong effect ($\eta_p{}^2$=0,434). Referring to item b) the results were not as expected and showed no effect. However, it is possible that the number of participants was too small for identifying an effect. Further the task als implies a high positive guessing rate.

Due to the small number of participants in the second pilot study, we only provide descriptive data. In Figure 11 we show the considerable increase of the correct responses for the performance task as well as the covariation task.

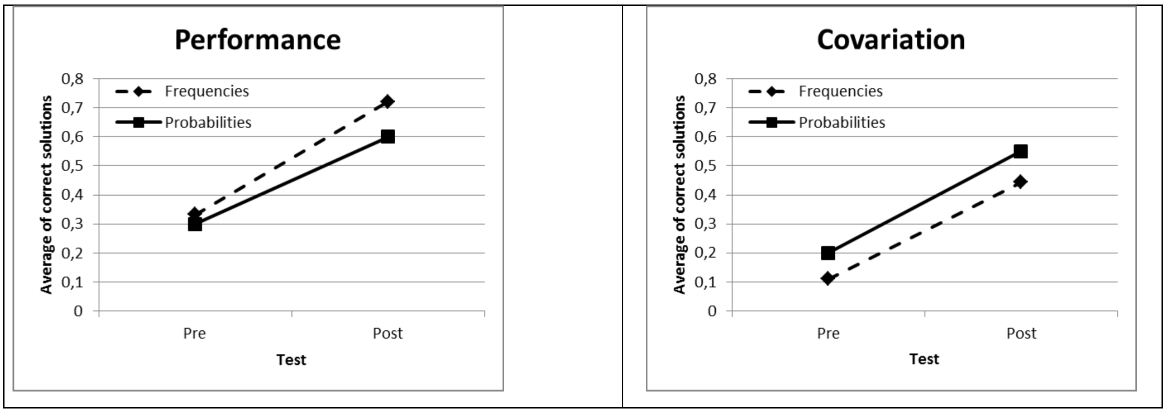


Figure 11. Descriptive results of the training concerning item a) and b)

Interestingly, there was an at most marginal difference between the frequency group and the probability group. Since, the participant were master students, there were some (but few) students that solved the task completely without any mistake. For this reason, the proportion of correct answers was initially about 30% referring to the performance task (item a). In the same way, the increase of ability to solve a task representing the covariation aspect is striking.

Besides the master students who were completely able to deal with Bayesian situations, the big shift of dealing with Bayesian situations is also striking in a qualitative way. Figure 12 shows exemplarily the solution of one of the students. In the pre-test, this student tried to apply the tree diagram to solve the problem but failed without presenting a solution. After the training, the unit square was applied without any mistake.

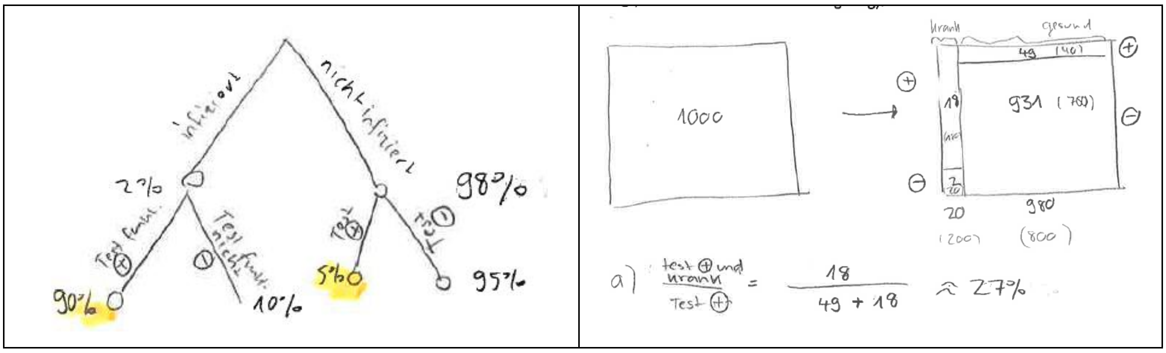


Figure 12. Result in the pre-test and the post-test

Finally, although most of the students used the tree diagram in the pre-test, the unit square seems to be convincing for the students since most of the students used the unit square in the post-test (Table 2; few students used two different visualizations).

Table 2. Number solutions with a specific visualization (two items, 19 students)

	Pre-test	Post-test
Unit square	0	30
Tree diagram	17	2
2x2-table	18	2
Only symbols	4	0
No visualization	7	5

6. Discussion and conclusion

Although dealing with Bayesian situations is obviously difficult, there are powerful strategies to increase people's performance as well as people's ability to deal with problems representing the covariation aspect.

Our research firstly yielded that visualization increase the people's ability of dealing with Bayesian situations. Particularly, our research yielded strong evidence that a visualization that makes the nested-sets structure of a Bayesian situation transparent is efficient. This result could inform the educational practice in which the tree diagram is the most common visualization when dealing with Bayesian situations. However, regarding our research results the tree diagram could be replaced by other visualizations, e.g. the double tree diagram or the unit square.

As expected, a training result is an increase of the ability of dealing with Bayesian situations. This is the case referring to performance, but also referring to the covariation aspect. Interestingly, also a very brief training lasting only 30 minutes show a considerable effect: From about 9% in the pre-test to about 80% in the group of students of grade 11. A further interesting question is whether the positive effect is also apparent in a follow up test after some weeks.

Our second pilot study yielded the result of a missing difference between the group using the format of natural frequencies and the group using the format of probabilities which is in contrast to former research results (e.g. Binder et al., 2015). However, since an understanding of Bayesian situations could include the ability to deal with parameter changes and probabilities behind solving performance tasks, the unit square could be an appropriate strategy to successfully facilitate a shift from natural frequencies to probabilities.

Finally, since we defined two visualizations as adequate for understanding Bayesian situations, further studies should focus also on double trees and the comparison of double trees and unit squares.

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. 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*, 241. https://doi.org/10.3389/fpsyg.2016.02026

Böcherer-Linder, K., & Eichler, A. (in review). The effect of visualizing statistical information for judgment and decision making in Bayesian situations. *Frontiers in Psychology*.

Böcherer-Linder, K., Eichler, A., & Vogel, M. (2017). The impact of visualization on flexible Bayesian reasoning. *Avances de Investigación en Educación Matemática - AEIM*, *11*, 25–46.

Böcherer-Linder, K., Eichler, A., & Vogel, M. (2018[a]). Visualising conditional probabilities - Three perspectives on unit squares and tree diagrams. In C. Batanero & E. J. Chernoff (Eds.), *Teaching and Learning Stochastics* (pp. 73–88). Cham: Springer International Publishing.

Böcherer-Linder, K., Eichler, A., & Vogel, M. (2018b). Visualizing statistical information with unit squares. In A. M. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the 10th International Conference on Teaching Statistics*. Kyoto, Japan: IASE.

Borovcnik, M. (2012). *Conditional probability – a review of mathematical, philosophical, and educational perspectives*. Topic Study Group 11 "Teaching and Learning Probability" at ICME 12, Seoul. Retrieved from https://www.researchgate.net/publication/304495166_Conditional_Probability_-_a_Review_of_Mathematical_Philosophical_and_Educational_Perspectives

Brase, G. L. (2009). Pictorial representations in statistical reasoning. *Applied Cognitive Psychology*, *23*, 369–381. https://doi.org/10.1002/acp.1460

Budgett, S., Pfannkuch, M., & Franklin, C. (2016). Building conceptual understanding of probability models: Visualizing chance. In C. R. Hirsch (Ed.), *Annual Perspectives in Mathematics Education 2016* (pp. 37–49). Reston, VA: National Council of Teachers of Mathematics.

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

Diaz, C., & Batanero, C. (2009). University students' knowledge and biases in conditional probability reasoning. *International Electronical Journal of Mathematics Education*, *4*, 131–162.

Diaz, C., Batanero, C., & Contreras, J. M. (2010). Teaching independence and conditional probability. *Boletin de Estadistica e Investigacion Operativa*, *26*, 149–162.

Eichler, A., & Vogel, M. (2013). *Leitidee Daten und Zufall: Von konkreten Beispielen zur Didaktik der Stochastik* (2., akt. Aufl. 2013). Wiesbaden: Springer. Retrieved from http://dx.doi.org/10.1007/978-3-658-00118-6

Friendly, M. (1999). Extending mosaic displays: marginal, conditional, and partial views of categorical data. *Journal of Computational and Graphical Statistics*, *8*, 373–395. https://doi.org/10.1080/10618600.1999.10474820

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. https://doi.org/10.1016/j.socscimed.2013.01.034

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*, 684–704. https://doi.org/10.1037/0033-295X.102.4.684

Hoffrage, U., Hafenbrädl, S., & Bouquet, C. (2015). Natural frequencies facilitate diagnostic inferences of managers. *Frontiers in Psychology*, *6*, 642. https://doi.org/10.3389/fpsyg.2015.00642

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

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

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

Oldford, R. W. (2003). Probability, problems, and paradoxes pictured by eikosograms. Retrieved from

http://www.stats.uwaterloo.ca/~rwoldfor/papers/venn/eikosograms/examples/paper.pdf

Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*, 529–556. https://doi.org/10.1016/S0959-4752(01)00030-5

Satake, E., & Murray A., V. (2014). Teaching an application of Bayes' rule for legal decision-making: Measuring the strength of evidence. *Journal of Statistics Education*, *22*, 1–29.

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

Sirota, M., Kostovičová, L., & Vallée-Tourangeau, F. (2014). How to train your Bayesian: a problem-representation transfer rather than a format-representation shift explains training effects. *Quarterly Journal of Experimental Psychology*, *68*, 1–9. https://doi.org/10.1080/17470218.2014.972420

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

Steckelberg, A., Balgenorth, A., Berger, J., & Mühlhauser, I. (2004). Explaining computation of predictive values: 2 x 2 table versus frequency tree. A randomized controlled trial ISRCTN74278823. *BMC Medical Education*, *4*, 13. https://doi.org/10.1186/1472-6920-4-13

Talboy, A. N., & Schneider, S. L. (2017). Improving accuracy on bayesian inference problems using a brief tutorial. *Journal of Behavioral Decision Making*, *30*, 373–388. https://doi.org/10.1002/bdm.1949

Tsai, J., Miller, S., & Kirlik, A. (2011). Interactive visualizations to improve Bayesian reasoning. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*, 385–389. https://doi.org/10.1177/1071181311551079

Veaux, R. D. de, Velleman, P. F., & Bock, D. E. (2012). *Intro stats* (3. ed., international ed., technology update). Boston, MA.: Pearson/Addison-Wesley.