VISUALIZING STATISTICAL INFORMATION WITH UNIT SQUARES

Katharina Böcherer-Linder[1], Andreas Eichler[2] and Markus Vogel[3]
[1]University of Education Freiburg, Germany
[2]University of Kassel, Germany
[3]University of Education Heidelberg, Germany
katharina.boechererlinder@ph-freiburg.de

The representation of statistical information in the form of natural frequencies is one of the main factors that support Bayesian reasoning. However, materials for professionals in the fields of medicine or law usually present information in the form of probabilities. Therefore, one aim of our research is to develop a training for professionals in those fields. In this paper, we present a short intervention that uses unit squares as visualization of statistical information. We report on a pilot study where students at university were trained how translate the usual probability format to a frequency format and how to visualize the statistical information. The results encourage the approach to use unit squares as visualizations of statistical information for training Bayesian reasoning.

INTRODUCTION AND BACKGROUND

In many areas, such as medicine or law, professionals have to deal with risk. For example, doctors must interpret the results of diagnosis tests or lawyers must consider evidences appropriately. The Bayes rule is a mathematical model for the interpretation of the probability of a hypothesis in the light of new evidence. The following example illustrates such a situation in which the Bayes rule can be applied:

Table 1**.** Probability version of a Bayesian reasoning situation (traveler-example).

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 uninfected 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 had been infected?

The Bayes' rule allows calculating the posterior conditional probability of having the disease given a positive test result. For the situation of the example in Table 1 the result is:

$$P(disease|test\ positive) = \frac{80\%\ \cdot 10\%}{80\% \cdot 10\% + 15\% \cdot 90\%} \approx 37{,}2\%$$

Despite the great relevance of the Bayes rule for real life situations, psychological and educational research gained evidence that people often fail when applying the Bayes rule (Kahneman, Slovik & Tversky, 1982; Gigerenzer & Hoffrage, 1995; Díaz, Batanero & Contreras, 2010). Eddy (1982) for example reported that only 5 out of 100 physicians interpreted the result of a medical diagnosis test (similar to the problem of Table 1) correctly. However, misinterpretations in the field of medicine or law can have severe consequences (e.g. Gigerenzer, 2002).

Nonetheless, mainly two methods have been identified how to improve performance in Bayesian reasoning tasks, i.e. the presentation of the statistical information in form of natural frequencies and the visualization of the statistical information (e.g. McDowell & Jacobs, 2017). In Figure 1, we show the frequency representation of the traveler-example (Table 1) at the basis of a sample of 1000 people and the visualization with a unit square:

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

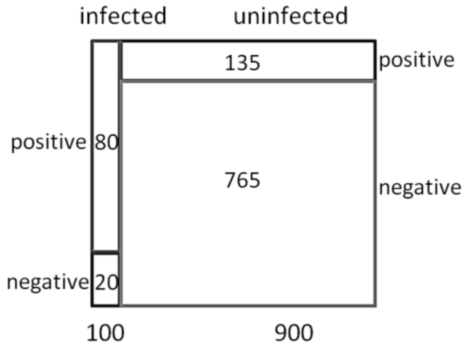
Frequency representation: 100 out of 1000 travelers contracted a new kind of disease during their trip. 80 out of 100 infected people get a positive test result. 135 out of 900 uninfected people get a negative test result.	

Figure 1. Frequency representation of the traveler-example and visualization of statistical information with the unit square.

When the statistical information is presented with natural frequencies, the posterior conditional probability of having the disease given a positive test result can be calculated as a simple proportion of the infected among the positively tested people:

$$P(disease|test\ positive) = \frac{80}{80+135} \approx 37{,}2\%$$

In general, when the statistical information is presented with natural frequencies as shown in Figure 1, performance increases substantially compared to problem representations with probabilities as shown in Table 1 (e.g. Gigerenzer & Hoffrage, 1995). The visualization of the statistical information can have an additional beneficial effect beyond the effect of natural frequencies (e.g. Garcia-Retamero & Hoffrage, 2013). However, the beneficial effect of the visualization depends on the specific form of the visualization and therefore turned out to be one of the main moderators of the frequency-effect (McDowell & Jacobs, 2017). Theoretically, the unit square was expected to be a suited visualization to endorse understanding of Bayesian reasoning problems (e.g. Eichler & Vogel, 2010; Oldford & Cherry, 2006). In an own study with 143 students the unit square outperformed the tree diagram, when the performance in Bayesian reasoning tasks was regarded (Böcherer-Linder & Eichler, 2017).

However, these two methods (the use of natural frequencies and the visualization of statistical information) lack ecological validity for professionals in the field of medicine or law, since the statistical information in the problems of daily practice in these fields are presented in probability versions similar to the example shown in Table 1. This leads to the need for training physicians or lawyers in solving Bayesian reasoning problems. However, only a few training studies have been carried out until now (e.g. Bea, 1995 or Sedlmeier & Gigerenzer, 2001), few with physicians (e.g. Steckelberg et al., 2004) and none of them with lawyers.

The aim of our research is to develop an effective training for professionals in fields that are only peripherally mathematics related, like medicine or law. In this paper, we present the results of a pilot study aiming to test variants of interventions for training Bayesian reasoning. We developed the training at the basis of the mentioned research results and present the results of a Pre-and Posttest with 15 students of mathematics education, before and after the training. We discuss the results with regard to former research results and with regard to our overall project of developing an effective training for professionals in the field of medicine or law.

THE TRAINING (20 MINUTES)

The training is based on two main research results of how to improve Bayesian reasoning:

- The use of natural frequencies
- The visualization of statistical information with unit squares

Since we took into account the ecological validity, we presented the problems in the probability version (c.f. Table 1) and trained how to translate the probability version into a frequency version and how to visualize the statistical information in the process of problem solving. The training had two phases: phase 1 (10 minutes) had three steps and was a short instruction in form of a worked example how to solve the problem of Table 1 with the help of the unit square. In Figure 2 and Figure 3, we show these three steps of the worked example of the training-phase 1 (translated from German):

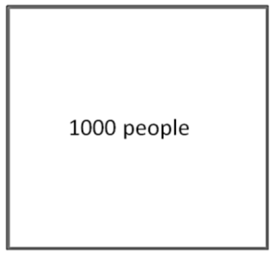
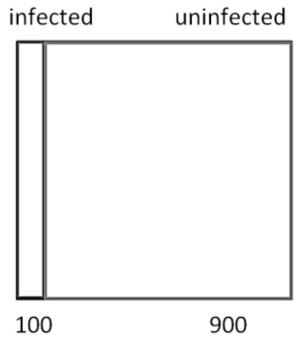
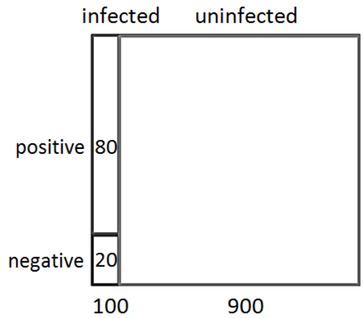
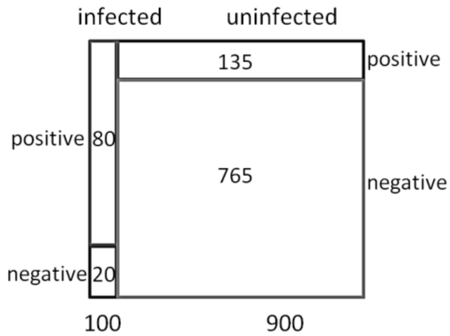
Problem presentation: see Table 1	
Step 1: Choice of the sample size	
For solving the problem, we first consider the question what the probabilities mentioned in the text imply for a concrete group of travelling people. We choose a sample size, for example 1000 people. We represent the group of 1000 people by drawing a unit square.	
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.	

Figure 2. Step 1 and 2 of the worked example in the training phase 1.

Step 1 and Step 2 shown in Figure 2 were an instruction how to construct a frequency representation with the help of the unit square. Step 3 (Figure 3) is an explanation how to extract the relevant numbers from the visualization and how to calculate the correct proportion that corresponds to the posterior conditional probability. Additionally, we visualized the requested proportion as a “picture-formula” (c.f. Eichler & Vogel, 2010):

Step 3: Calculating the correct proportion

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 we border 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%.

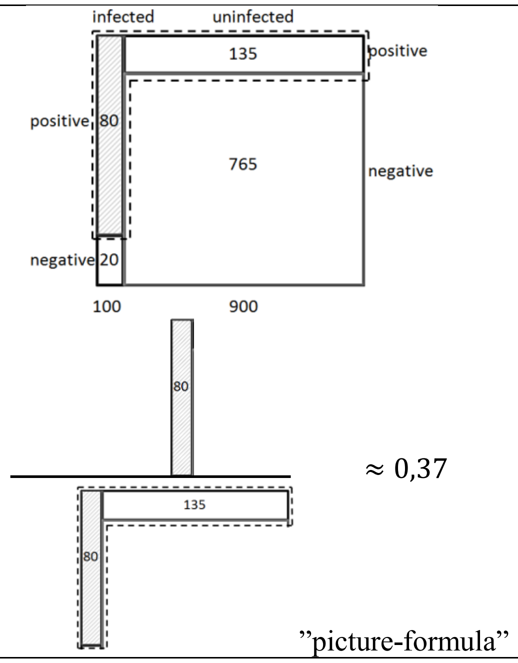


Figure 3. Step 3 of the worked example in the training phase 1.

Phase 2 (10 minutes) involved an exercise that was structurally identical to the worked example (Figure 2 and Figure 3) but had another context (Trisomy). An oral explanation and a presentation of the correct solution were given. Even though it was possible to draw the unit square true to scale in this case, we only showed a rough drawing of the unit square, since we wanted to enable the participants to work with the visualization as a tool for problem solving which did not necessarily need a precise drawing. Moreover, in more general situations, it would not always be possible to represent extremes values true to scale.

PRE- AND POSTTEST

Before and after the training, the participants (15 university students) answered three Bayesian reasoning problems, A, B and C, represented in the probability version (c.f. Table 1). These problems were structurally identical to that shown in Table 1 and the contexts were an A: economics problem (c.f. Binder et al., 2015), B: measles infection (c.f. Gigerenzer & Hoffrage, 1995) and C: spam filter (c.f. Figure 3). The answers were coded and anonymity was guaranteed.

RESULTS

For all of the three problems A, B and C, the number of correct answers increased substantially, from 8 overall correct answers in the pretest to 32 correct answers in the posttest (see Table 2):

Table 2. Number of correct answers in the pretest and the posttest.

Bayesian problem	Pretest	Posttest	Maximum
A (economics problem)	4	11	15
B (measles infection)	2	12	15
C (spam filter)	2	9	15

A more qualitative evaluation showed that many participants that were not able to solve the problems in the pretest, successfully constructed a frequency representation by drawing a unit square in the posttest and, thus, solved the problems correctly. In Figure 4, we show an example illustrating the effect of the training:

<table>
<tr><td rowspan="2">Problem C (spam filter):
10% of the emails to a company are spam-mails.
85% of the spam-mails are identified correctly by a spam filter.
2% of the non-spam-mails are erroneously identified as spam.

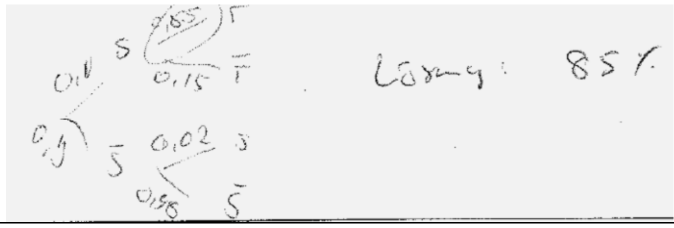
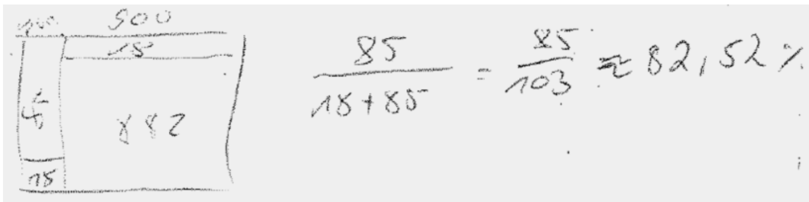
What is the probability that an email that is identified as spam-mail is actually spam?</td><td>Pretest:

</td></tr>
<tr><td>Posttest:

</td></tr>
</table>

Figure 4. Example of one student's answers in the Pretest and in the Posttest.

In the pretest, the student tried to find the solution by drawing a tree diagram with probabilities. The student encircled the "0,85" at the branch and wrote "85%" for solution. Apparently, the student confounded the conditional probabilities $P(A|B)$ and $P(B|A)$, which is known as a common error in the field of conditional probabilities (e.g. Díaz, Batanero & Contreras, 2010). After the training, the student did not make this fail any more. The student constructed a frequency representation by drawing the unit square, he was able to extract the relevant numbers and he calculated the requested probability correctly.

DISCUSSION

Although no control group was tested in this pilot study and hence it is not possible to exclude a learning effect from the pretest to the posttest, the results suggest a positive effect of the training. This is all the more the case, since control groups in other studies with similar problems (e.g. Sedlmeier & Gigerenzer, 2001; Sirota, Kostovičová & Vallée-Tourangeau, 2015) showed hardly any learning effect. Moreover, the considerable effect of the training with the unit square was reached in only 20 minutes which is much shorter than the training of Sedlmeier and Gigerenzer (2001).

It remains an open question, if similar effects could be reached by using other visualizations like tree diagrams or 2x2-tables for training the construction of frequency representation. Sedlmeier and Gigerenzer (2001) for instance used tree diagrams and frequency grids for "visualization training" and observed beneficial effects as compared to "rule training" and "no training". However, several participants in our pilot study used tree diagrams or 2x2-tables in the pretest (see Figure 4 for an example) and did not find the correct solution.

The conclusion is that the unit square seems to be an effective visualization for frequency representation construction and that it is a promising approach to use unit squares as visualizations of statistical information for training Bayesian reasoning.

REFERENCES

Bea, W. (1995). *Stochastisches Denken [Stochastical Reasoning]*. Frankfurt: Lang.

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), 1–9. DOI: 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, 8*(2026), 1–11. DOI: 10.3389/fpsyg.2016.02026.

Díaz, C., Batanero, C. & Contreras, J., M. (2010). Teaching independence and conditional probability. *Boletín de Estadística e Investigación Operativa, 26*(2), 149-162.

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

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

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. DOI: 10.1016/j.socscimed.2013.01.034.

Gigerenzer, G. (2002). *Calculated Risks: How to Know When Numbers Deceive You*. New York, Simon & Schuster.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction. Frequency formats. *Psychological Review, 102(4)*, 684–704. DOI: 10.1037/0033-295X.102.4.684.

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*, *143(12)*, 1273–1312. DOI: 10.1037/bul0000126.

Oldford, R. W., & Cherry, W. H. (2006). Picturing probability: The poverty of Venn diagrams, the richness of eikosograms. Online: http://www.stats.uwaterloo.ca/~rwoldfor/research.html

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

Sirota, M., Kostovičová, L., & Vallée-Tourangeau, F. (2015). Now you Bayes, now you don’t : effects of set-problem and frequency-format mental representations on statistical reasoning. *Psychon. Bull. Rev.*, *22*, 1465-1473. DOI: 10.3758/s13423-015-0810-y.

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.