The impact of visualization on flexible Bayesian reasoning

Katharina Böcherer-Linder[1], University of Education Freiburg (Germany)

Andreas Eichler[2], University of Kassel (Germany)

Markus Vogel[3], University of Education Heidelberg (Germany)

Recibido el 20 de julio de 2016; aceptado el 16 de enero de 2017

El impacto de la visualización sobre el razonamiento Bayesiano flexible

Resumen

Hay un amplio consenso en que la visualización de la información estadística puede apoyar el razonamiento bayesiano. El trabajo se enfoca en la comprensión conceptual de las situaciones que implican razonamiento bayesiano e investiga si el diagrama en árbol o el cuadrado unidad son más apropiados para apoyar la comprensión de la influencia de las tasas base, que son introducidas como parte flexible del razonamiento bayesiano. Como gráfico estadístico, el cuadrado unidad refleja la influencia de las tasas base no sólo de forma numérica, sino también geométrica. En consecuencia, en dos experimentos con estudiantes de grado (N = 148 y N = 143) se obtuvieron mejores resultados con el cuadrado unidad que con el diagrama en árbol para comprender la influencia de las tasas base. Nuestros resultados contribuyen a la discusión sobre cómo visualizar las situaciones bayesianas y tiene consecuencias prácticas para la enseñanza y el aprendizaje de la estadística.

Palabras clave. *Razonamiento bayesiano, tasas base, visualización, diagrama en árbol, cuadrado unidad.*

O impacto da visualização no raciocínio Bayesiano flexível

Resumo

É amplamente consensual que a visualização da informação estatística pode apoiar o raciocínio bayesiano. Este artigo foca-se na compreensão conceptual de situações que envolvem raciocínio Bayesiano e investiga se o diagrama de árvore ou o quadrado unitário é mais apropriado para apoiar a compreensão da influência da taxa de base que é introduzida como parte do raciocínio bayesiano flexível. Como gráfico estatístico, o quadrado unitário reflete a influência da taxa de base não apenas de forma numérica mas também geométrica. Consequentemente, em duas experiências com estudantes universitários (N = 148 e N = 143), obtiveram-se melhores desempenhos com o quadrado unitário do que com o diagrama em árvore, no que se refere à compreensão da influência da taxa de base. Os nossos resultados podem informar a discussão sobre como visualizar situações bayesianas e têm consequências práticas para o ensino e aprendizagem da estatística.

Palavras-chave. Raciocínio bayesiano, taxa de base, visualização, diagrama de árvore, quadrado unitário.

The impact of visualization on flexible Bayesian reasoning

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

Abstract

There is wide consensus that visualizations of statistical information can support Bayesian reasoning. This article focusses on the conceptual understanding of Bayesian reasoning situations and investigates whether the tree diagram or the unit square is more appropriate to support the understanding of the influence of the base rate, which is introduced as being a part of flexible Bayesian reasoning. As a statistical graph, the unit square reflects the influence of the base rate not only in a numerical but also in a geometrical way. Accordingly, in two experiments with undergraduate students (N = 148 and N = 143) the unit square outperformed the tree diagram referring to the understanding of the influence of the base rate. Our results could inform the discussion about how to visualize Bayesian situations and has practical consequences for the teaching and learning of statistics.

Key words. Bayesian reasoning, base rate, visualization, tree diagram, unit square.

L'influence de la visualisation sur le raisonnement bayésien flexible

Résumé

Selon un large consensus, la visualisation de l'information statistique peut aider le raisonnement bayésien. Cet article porte sur la compréhension conceptuelle des situations qui impliquent le raisonnement bayésien et il cherche à déterminer si le diagramme arborescent ou le carré d'unité est plus approprié pour aider à la compréhension de l'influence de la fréquence de base, qui est introduite comme étant une partie du raisonnement bayésien flexible. En tant que graphique statistique, le carré d'unité reflète l'influence de la fréquence de base pas seulement de manière numérique mais aussi de manière géométrique. Par conséquent, dans deux expériences avec des étudiants (N = 148 et N = 143) le carré d'unité a été plus efficace que le diagramme arborescent pour la compréhension de l'influence de la fréquence de base. Nos résultats contribuent à la discussion sur la visualisation des situations bayésiennes et ont des conséquences pratiques pour l'enseignement et l'apprentissage des statistiques.

Mots-clés. Raisonnement bayésien, fréquence de base, visualisation, diagramme arborescent, carré d'unité.

1. Introduction

As a part of risk literacy, decision making under uncertainty has been estimated to be an important subject of mathematics education (Spiegelhalter & Gage, 2014). One reason is the great relevance for real life situations: For example, crucial decisions under uncertainty can depend on the interpretation of the outcomes of medical diagnosis tests, DNA tests or other diagnostic tests. However, the interpretation of the outcomes of a medical diagnosis test and thus decision making under uncertainty is susceptible to errors and misunderstanding (e.g. Ellis, Cokely, Ghazal, & Garcia-Retamero, 2014; Gigerenzer & Hoffrage, 1995). For example, if the base rate, i.e. the proportion of incidences in a population, is very low, a counterintuitive statistical problem situation emerges where false positive test results are more probable than true positive test results. In such situations people tend to ignore the influence of the base rate which is known as the "base rate neglect" (Kahneman & Tversky, 1996).

In recent decades, research gained some knowledge about how to promote insight into Bayesian reasoning situations. One approach that has extensively been studied is to represent the information in form of natural frequencies instead of probabilities (e.g. Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995; Hoffrage, Gigerenzer, Krauss, & Martignon, 2002). A second approach is to focus on different kinds of visualizing the statistical information (e.g. Binder, Krauss, & Bruckmaier, 2015). The cited studies investigated the effects of using natural frequencies or visualizations on calculating posterior probabilities according to the Bayes' rule.

In our research approach, we focus on investigating Bayesian thinking beyond calculating probabilities and on studying conceptual understanding of Bayesian situations. Aiming to identify a best teaching method for conditional probabilities in the field of Bayesian reasoning, Borovcnik (2012) propose some approaches including "to investigate the influence of variations of input parameters on the result" (p. 21). In this regard, we use the term "flexible Bayesian reasoning" and define it as the understanding of parameter dependency in Bayesian reasoning situations. For example, one parameter that strongly impacts the result of Bayes' rule is the base rate. As it has been shown that – when using natural frequencies – visualizations can support Bayesian reasoning (e.g. Binder et al., 2015; Brase, 2009, 2014; Sedlmeier & Gigerenzer, 2001) we discuss the benefit of visualizations with natural frequencies for flexible Bayesian reasoning. Especially, our aim is to identify a visualization of statistical information that is helpful for the estimation of the influence of the base rate. In this context, flexible Bayesian reasoning could be understood as being a part of "reading beyond the data" which is the highest category of graph comprehension (Curcio, 1989).

In this article, we compare the effects of two visualizations, i.e. the tree diagram and the unit square on flexible Bayesian reasoning. For this, we firstly focus on the effect that different base rates can have. Further, we briefly outline the results of psychological research concerning the base rate neglect afterwards. Subsequently, we show how statistical information and the influence of the base rate on conditional probabilities can be visualized. In two experiments, we compare the effects of the unit square and the tree diagram both based on natural frequencies for flexible Bayesian reasoning. Finally, we discuss didactical implications based on our empirical evidence.

2. The impact of the base rate on conditional probabilities

We discuss the impact of the base rate on conditional probabilities by using the example of a HIV diagnosis test. Assume that a test for detecting HIV has the following characteristic (this is approximately the characteristic of the ELISA test; e.g. a German institution that supports people who are seeking for help: www.dresden.aidshilfe.de):

- 99.5% of infected people get a positive test result (sensitivity).
- 0.5% of uninfected people get a positive test result (which corresponds to a specificity of 99.5%).

Because of the high sensitivity and even high specificity, people would generally expect a high accuracy of the HIV-test. But this is not always the case! The interpretation of a positive test result strongly depends on the base rate. Imagine, running this HIV-test in a population where the base rate, i.e. the proportion of infected people in the population, is high, such as in Malawi where the base rate is approximately 10% at present time (www.unaids.org). For this population, the posterior probability of being infected given a positive test result is according to the Bayes' rule:

$$P(infected|positive) = \frac{10\% \cdot 99.5\%}{10\% \cdot 99.5\% + 90\% \cdot 0.5\%} \approx 96\%$$

However, if the same test is running in a population with a low base rate, for example in Germany, where the base rate is 0.1% at present time (www.unaids.org), the probability of being infected given a positive test result is only 17% according to the Bayes' rule:

$$P(infected|positive) = \frac{0.1\% \cdot 99.5\%}{0.1\% \cdot 99.5\% + 99.9\% \cdot 0.5\%} \approx 17\%$$

Since people tend to ignore the influence of the base rate (Bar-Hillel, 1980; Gigerenzer & Hoffrage, 1995; Tversky & Kahneman, 1974), the result of only 17% is counterintuitive with regard to the high sensitivity (99.5%) and high specificity (99.5%). This counterintuitive phenomenon occurs when the conjoint probability of infected and positively tested (0.1% · 99.5%) is smaller than the conjoint probability of uninfected and positively tested (99.9% · 0.5%).

In our research, we are interested in identifying a visualization of the statistical information that helps to mentally represent such situations beyond merely calculating the probabilities and comparing them. This is in line with Borovcnik's claim (2012) that "more investigation and deliberations" (p. 21) are needed, especially "to clarify the structure of the situation, e.g. why a conditional probability is so small" (p. 21). Thus, for flexible Bayesian reasoning, i.e. understanding the influence of the base rate, we plead for a visualization that reflects the structure of the situation. For the design of such visualizations we take into account the following psychological research results.

3. Base rate neglect in psychological research

Psychological research gained evidence, that people struggle when dealing with conditional probabilities: Initiated by the work of Kahneman and Tversky in the 1970's, the "heuristics and biases" program studied human judgement under conditions of uncertainty. They primarily asked people to make judgements of single-event probabilities and found that human intuition often seems to be misleading (Tversky & Kahneman, 1974). For example, for the estimation of posterior probabilities, dramatic deviations from the correct values according to the Bayes' rule have been observed. To explain this deviation Kahneman and Tversky propose that people use heuristics that are generally useful in simplifying information but may result in systematic biases. One phenomenon that has been extensively documented is the base rate neglect, which refers to the tendency for people to ignore or underweight base rate in Bayesian reasoning problems when posterior probabilities have to be determined (Bar-Hillel, 1980; Gigerenzer & Hoffrage, 1995; Kahneman & Tversky, 1996). This bias has been observed and discussed also in educational research (e.g. Díaz, Batanero, & Contreras, 2010).

An influential explanation for the base rate neglect was suggested by Cosmides and Tooby (1996) and Gigerenzer (1994) departing from an evolutionary point of view: Since in our environment single-event probabilities are not observable, our mind has not evolved to process single-event probabilities. Therefore experimental materials using single-event probabilities lack ecological validity. Instead of presenting the information as probabilities they propose to use natural frequencies since "the mind is tuned to frequency formats, which is the information format humans encountered long before the advent of probability theory" (Gigerenzer & Hoffrage, 1995, p. 697). Indeed, experimental research gave evidence that presenting statistical information in the external format of natural frequencies, can improve Bayesian reasoning dramatically (e.g. Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995).

What does this concretely mean for the mentioned example of HIV test accuracy? In Figure 1, we show that the situation of the HIV diagnosis test becomes much clearer when we use natural frequencies instead of probabilities to represent the statistical information:

Probabilities:

In Germany, the probability that a person is infected with HIV is 0.1%. If the person is infected, the probability of getting a positive test result is 99.5%. If the person is not infected, the probability of getting a positive test result is 0.5%.

Natural frequencies:

In Germany, 1000 out of 1 000 000 people are infected with HIV. Of these 1000 infected people, 995 get a positive test result. Of the 999 000 uninfected, 4995 get a positive test result.

Figure 1. Two ways of representing the same statistical information.

When using the information format of probabilities the influence of the base rate is in some sense hidden: It is not immediately clear that the accuracy of a positive test result is low because the base rate is low. In contrast, when the same information is represented in form of natural frequencies the influence of the base rate becomes transparent since it is clear that the false positive rate (represented by the number of 4995 people) is much more than the true positive rate (represented by the number of 995 people). Thus, people would no longer expect that the probability of being infected given a positive test result is very high when running the test in Germany because much more people are expected to get a positive test result although not being actually infected. Since our study is concerned with the impact of visualization on the understanding of the influence of the base rate we will take into account the facilitating effect of natural frequencies for the design of the visualizations.

4. Visualizing Bayesian situations: The unit square and the tree diagram

Under the assumption that the information format of natural frequencies facilitates Bayesian reasoning (Gigerenzer & Hoffrage, 1995), we use natural frequencies for both, the tree diagram and the unit square (see Figure 2). The unit square was often used with probabilities in research literature and educational literature (e.g. Bea, 1995; Eichler & Vogel, 2010; Oldford, 2003; Sturm & Eichler, 2014), but we changed the common unit square with probabilities into a unit square with frequencies. In the unit square, the area of the whole square represents the total sample size (see Figure 2). We insert the numbers into the rectangular partial areas which are proportional to the quantities they are meant to visualize (see Figure 2). In this way, we obtain a visualization of the base rate as the width of the rectangles on the right and the true positive rate (sensitivity), the false positive rate, the true negative rate (specificity) and the false negative rate as the height of the different rectangles. Finally the areas of the rectangles represent the conjoint events, for example the event of being infected and positively tested. Recent research gave evidence that the unit square with natural frequencies is an effective visualization to improve performance in the calculation of the Bayes' rule (Böcherer-Linder & Eichler, in press).

We compare the effect of the unit square with the well-known tree diagram as it has been used for example in the work of Sedlmeier and Gigerenzer (2001). The tree diagram with natural frequencies is a common visualization in probability education and risk communication (Kurz-Milcke, Gigerenzer & Martignon, 2008; Spiegelhalter & Gage, 2014; Veaux, Velleman, & Bock, 2012). Scholars often use tree diagrams to visualize Bayesian reasoning situations (Gigerenzer & Hoffrage, 1995; Mandel, 2014; Navarrete, Correia & Froimovitch, 2014). Although the effectiveness of both the tree diagram and the unit square was found in different studies (e.g. Bea, 1995; Sedlmeier & Gigerenzer, 2001) there is at the moment no study available – at least to our best knowledge – that compare different visualizations concerning their facilitating effect on flexible Bayesian reasoning.

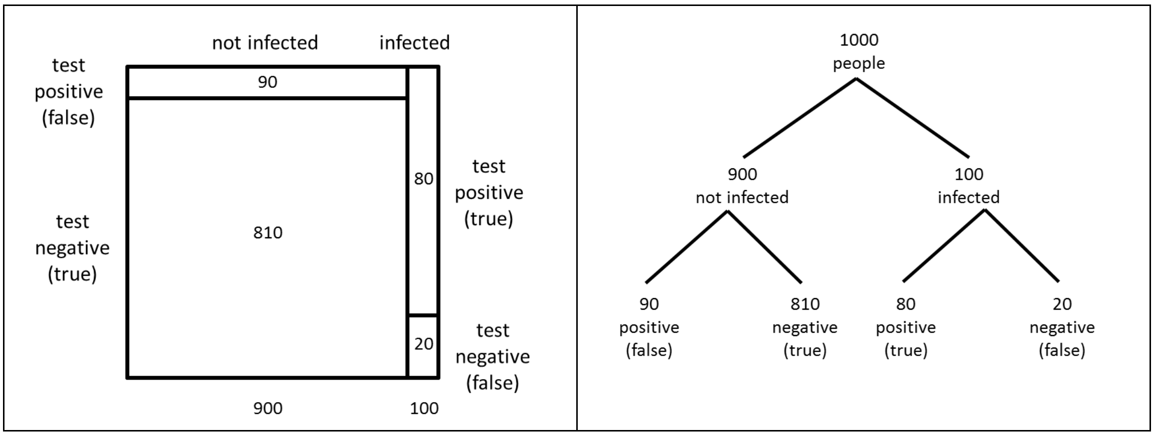


Figure 2. Example of the unit square and the tree diagram with natural frequencies for a base rate of 10%, a sensitivity of 80% and a specificity of 90%.

The unit square and the tree diagram bear the same numerical information (see Figure 2). However, they have a quite different structure. The unit square is a statistical graph (Tufte, 2013), which means, that the sizes of the partitioned areas are proportional to the sizes of the represented data. Therefore the proportion of incidences in a population, i.e. the base rate, is represented in a numerical and geometrical sense. Imagine a medical diagnosis test (e.g. sensitivity of 80% and specificity of 90%) running for three different populations with different base rates. As shown in Figure 3, the "visual appearance" of the unit square is changing: when the base rate increases, the vertical line moves to the left. In contrast, the tree diagram is not a statistical graph because the data are merely represented by numbers. Changing base rates only produce changing numbers, the "visual appearance" of the tree however does not change (see Figure 3).

Consequently, the unit square and the tree diagram encourage different mental representations of the data. Even if we had drawn only the first unit square in Figure 3 for the base rate of 10% the reader would be able to imagine the shift of the vertical line when the base rate increases (or decreases) and therefore could imagine the structure of the data for higher (or lower) base rates. In contrast, the tree diagram encourages a mental representation where the change of base rates is represented by greater or smaller numbers and their relations. In the same way, a unit square could encourage a mental representation of the influence of a base rate change on every proportion that is given in the unit square. For example, the unit square could visualize the change of the proportion of being infected given a positive test result if the base rate changes (Figure 3). Thus, we hypothesize that the unit square is appropriate to promote flexible Bayesian reasoning.

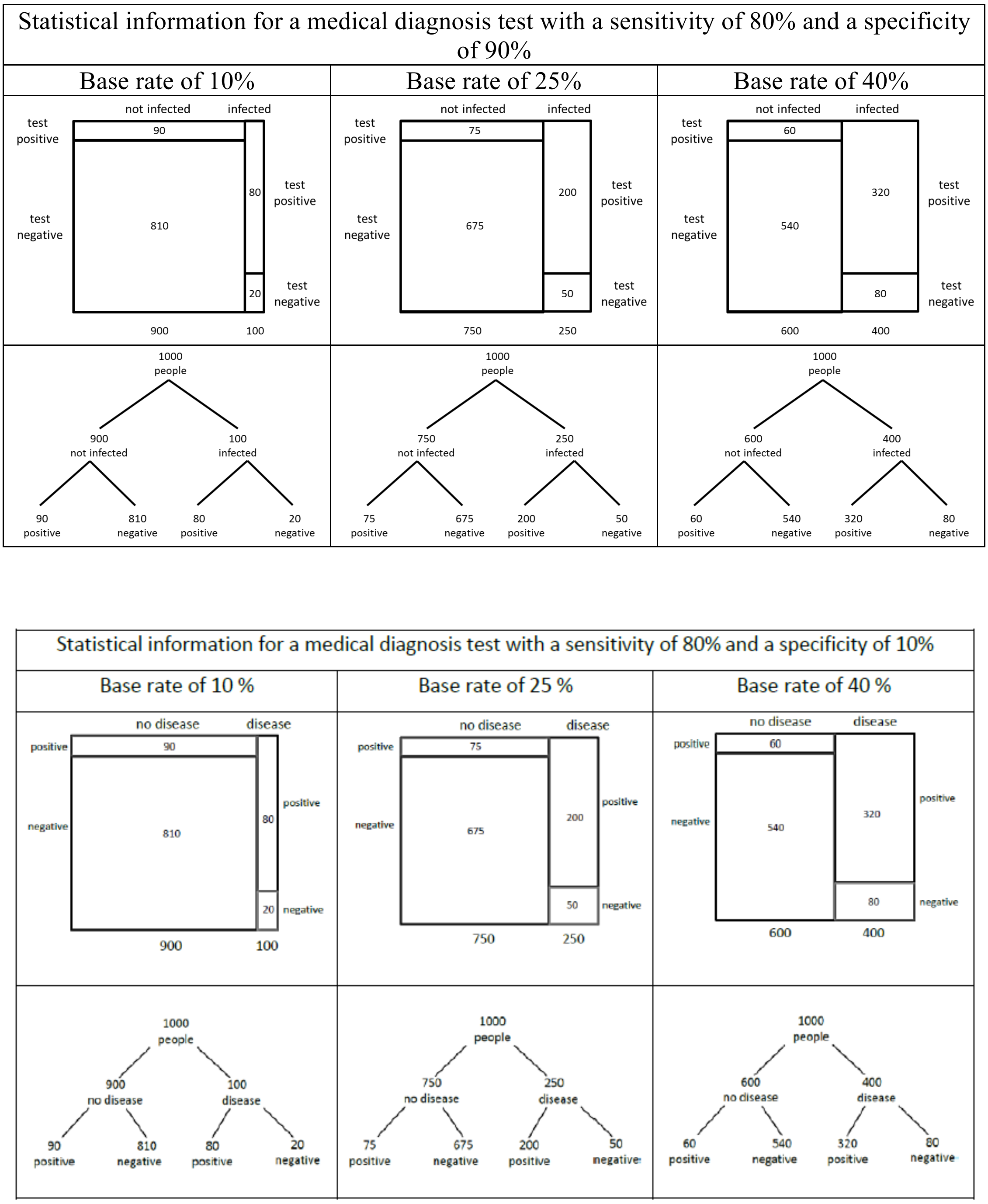


Figure 3. Visualizations of changing base rates.

5. Research question

In the previous sections, we pointed out the benefit of two visualizations for understanding the influence of the base rate in Bayesian reasoning situations. Our analysis of the structure of the unit square suggests that the unit square is appropriate to visualize the influence of the base rate. We compare the unit square with the tree diagram, since the tree diagram with natural frequencies is a common visualization in probability education (Spiegelhalter & Gage, 2014) and its efficiency in Bayesian reasoning situations has been proved (Sedlmeier & Gigerenzer, 2001). Thus, our research goal is to study whether the tree diagram or the unit square is more efficient to support flexible Bayesian reasoning.

To investigate this research question, we conducted two experiments with undergraduate students. We designed a questionnaire where we asked to estimate the effect of changing base rates for different proportions in the data. As shown in section 4, the unit square visualizes the influence of the base rate geometrically and numerically, whereas the tree diagram does it only in a numerical way. Thus, we hypothesized that the unit square is more efficient than the tree diagram to promote the understanding of the influence of the base rate on proportions in the data and therefore to support flexible Bayesian reasoning.

6. Experimental studies

For both studies, the method was similar: We had two questionnaires, one showing unit squares, the other showing tree diagrams to present the statistical information. The tasks, the context stories and the numerical information were identical, only the visualization differed. The participants got a brief description of a visualization based on a simple example that was shown at the beginning of the questionnaire. Those participants who got the questionnaire with the unit square got the description of the unit square, those who got the questionnaire with the tree diagram got the description of the tree diagram. In a previous study, where we used the same descriptions to introduce the visualizations, we could show that participants do not differ in extracting relevant numbers from the unit square and the tree diagram, both based on natural frequencies (Böcherer-Linder & Eichler, 2015). Thus, we can exclude any bias resulting from the reading of simple information.

6.1. Study 1

Sample and method: The participants were 148 undergraduates at the Heidelberg University of Education (Germany). They were beyond their first semester of their study and were enrolled in a course of mathematics education. In this course, the two visualizations and the Bayes' rule were not part of the curriculum. The participants were randomly assigned to a unit square group (N = 74) and to a tree diagram group (N = 74).To investigate the effect of visualization on flexible Bayesian reasoning, we designed test items where the participants have had to choose whether proportions in the data will become bigger, smaller or remain constant, when the base rate increases. In the Appendix, we provide the questionnaire of study 1 that shows unit squares. Note, that we only showed one visualization for each situation and asked what would happen if the base rate became bigger. Thus, participants had to imagine what would change and how the numerical values and the visual appearance of the visualizations would change. But we did not control if the participants really imagined the situation. We only rated if the answer was correct (1) or incorrect (0).

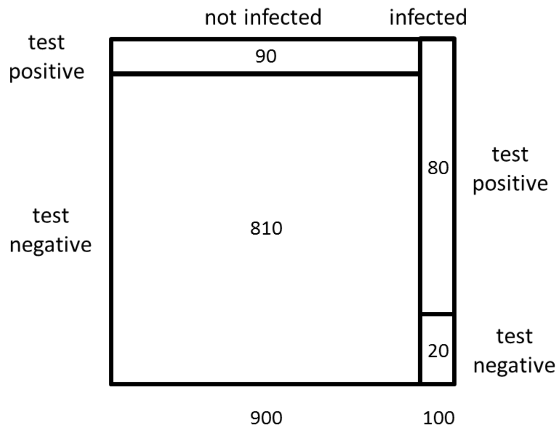
Altogether, we had three contextually different but mathematically identical situations concerning flexible Bayesian reasoning (1. smoke, 2. diagnosis and 3. flowers) and for each of the three situations four structurally identical items (a), (b), (c) and (d) (see Appendix). In Figure 4, we explain in more detail the structure of the test items using the second situation "diagnosis" as an example. In the small visualizations, we indicate the changes of the base rates by arrows. The shape of the tree is not influenced by a change of the base rate, thus there are no arrows in the trees. The shaded areas in the small visualizations in Figure 4 correspond to the numerators of the proportions that are addressed in the items. The denominators are marked by dotted lines. All items with number (a) in the questionnaire addressed proportions that remain constant when the base rate increases. All items with number (b) and with number (d) addressed proportions that increase with increasing base rates and that correspond to posterior conditional probabilities. All items with number (c) addressed proportions that decreased with increasing base rates and of which the denominators were equal to the total sample size.

Original test item:

Diagnosis (Study 1)

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:

How the following proportions change if, in the next routine screening test with 1000 people, the percentage of people infected with the disease is bigger?



Mark the correct solutions.

a. The proportion of people tested positive among the infected will be bigger / smaller / constant.
b. The proportion of infected among the people tested positive will be bigger / smaller / constant.
c. The proportion of people tested negative among all tested people will be bigger / smaller / constant.
d. The proportion of infected among the people tested negative will be bigger / smaller / constant.

Explanation of the test item:

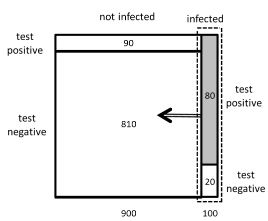
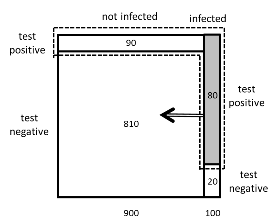
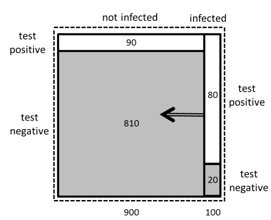
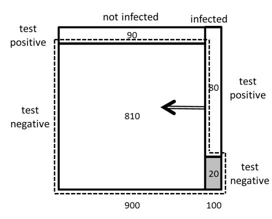
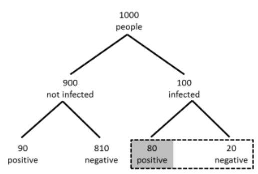
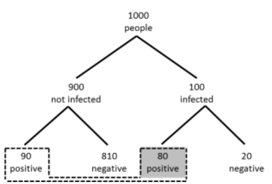
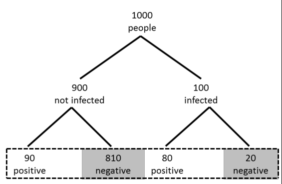
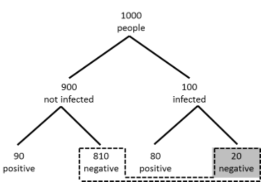
Item (a)	Item (b)	Item (c)	Item (d)
The proportion corresponding to $P(positive \mid infected)$ is constant if the base rate increases	The proportion corresponding to $P(infected \mid positive)$ increases if the base rate increases	The proportion corresponding to $P(negative)$ decreases if the base rate increases	The proportion corresponding to $P(infected \mid negative)$ increases if the base rate increases
			
			

Figure 4. For the diagnosis situation (see Appendix), we highlight the proportions that are addressed by items (a) to (d) in the corresoponding unit squares and tree diagrams.

Results: Since we rated each item with 1 (correct) or 0 (incorrect) and we had three situations with 4 items each, the maximum value for the questionnaire was 12 points. The 3 x 4 items of the questionnaire showed an appropriate reliability (α = .750). For the sum of the twelve items a t-test yielded the following result: The unit square (M = 6.58, SD = 3.277) was more efficient than the tree diagram (M = 5.57, SD = 2.700), t (140.843) = 2.053, p = .042, with an effect size of d = .34. Thus, our hypothesis that the unit square is more efficient to support flexible Bayesian reasoning was confirmed.

Although the effect size indicated a small effect, a qualitative analysis of the data showed that this effect was stable regarding different aspects: There are two different ways how to look at the data in a more detailed way: First, one can focus on the different kinds of proportions addressed by the different items. Since all items with number (a) addressed the same kind of proportion, we denote by "items_a" the sum of smoke_a + diagnosis_a + flowers_a (see Appendix). Thus the maximum value for items_a is 3 (and similarly for items_b, items_c and items_d). In Figure 5a, we provide the results for the different kinds of proportions represented in both the tree diagram and the unit square. Error bars represent the 95% CI of the means. Standard deviations for the tree diagram: items a (SD = 1.122), items_b (SD = 0.946), items_c (SD = 1.010), items_d (1.037). Standard deviations for the unit square: items_a (SD = 1.138), items_b (SD = 0.995), items_c (SD = 1.076), items_d (SD = 1.049).Figure 5a shows for all kinds of proportions the tendency that there is an advantage of the unit square. The results for the items (b) are for both visualizations a little bit higher than the other results. We assume that this phenomenon might be due to a learning effect from the items (a), since the proportion addressed in items (b) is the transposed of those addressed in items (a).

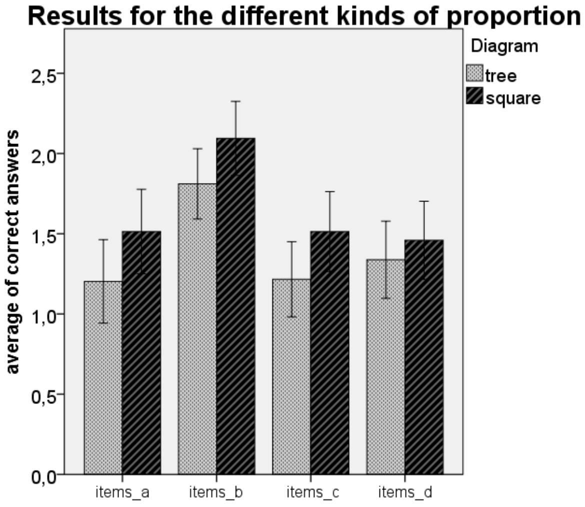


Figure 5a. Results in study 1 for the different kinds of proportion addressed by the different items.

Second, one can focus on the results for the different Bayesian situations "smoke", "diagnosis" and "flowers" (see Appendix) and summarize over the items for each situation (see Figure 5b). Error bars represent the 95% CI of the means. Standard deviations for the tree diagram: smoke (*SD* = 1.203), diagnosis (*SD* = 1.124), flowers (*SD* = 1.475). Standard deviations for the unit square: smoke (*SD* = 1.285), diagnosis (*SD* = 1.303), flowers (*SD* = 1.450). Thus, we denote by "smoke" the sum smoke_a + smoke_b + smoke_c + smoke_d. (and similarly for the other situations). Here, the maximum value for each situation is 4. Figure 5b shows that the tendency was an advantage of the unit square for this second aspect as well. The results for the situation "smoke" are for both visualizations a little bit higher than the results for the other situations. One plausible explanation is that the context of "smoke" is more adapted to the living environment of young people (cf. Siegrist & Keller, 2011). Nevertheless, the benefit of the unit square was not dependent on the problem context since we find an advantage of the unit square for all of the three different situations.

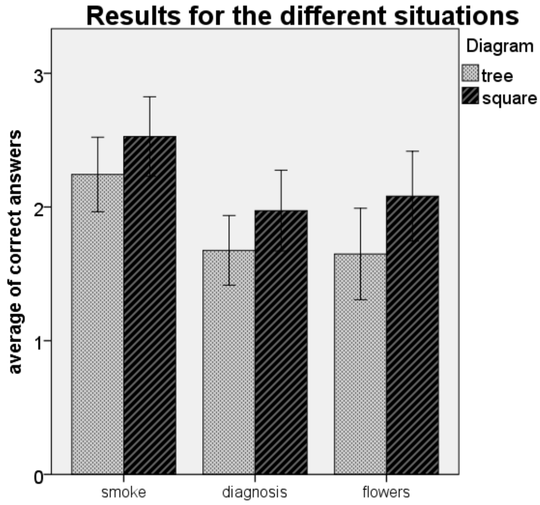


Figure 5b. Results in study 1 for the different Bayesian situations.

6.2. Study 2

Sample and method: The participants were 143 undergraduates at the Technical University of Munich (Germany). They were in the fourth semester of their study and were enrolled in a course of Electrical Engineering. In this course, the two visualizations and the Bayes' rule were not part of the curriculum. The participants were randomly assigned to the unit square (N = 74) and to the tree diagram (N = 69).

In this second study, we sought to replicate the results of study 1 with a focus on conditional probabilities. For this reason, we omitted all items with number (c) in the three situations "smoke", "diagnosis" and "flowers" used in the study 1 (see Appendix), because they did not refer to conditional probabilities. The rest of the items remained with little modifications (e.g. in the diagnosis situation we replaced in the first item the "people tested positive" by the "people tested negative", the item (b) remained identical and in the last item we replaced the "infected" by the "uninfected"). We did this modification in order to have each of the possibilities (bigger / smaller / constant) as an answer in the items. This little modification does not change the structure of the questions and therefore study 1 and study 2 are directly comparable. In Figure 6, we explain the structure of the test items using the second situation "diagnosis" as an example. Similarly to our explanation of the first experiment, the shaded areas in the small visualizations in Figure 6 highlight the numerators of the proportions that are addressed in the items. The denomiators are marked by dotted lines.

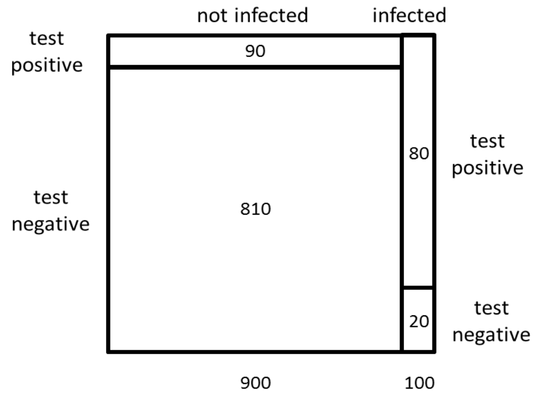
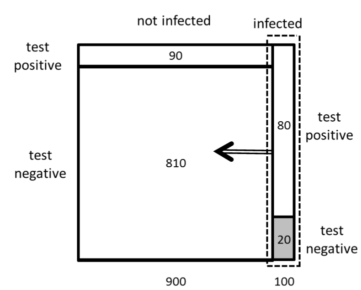
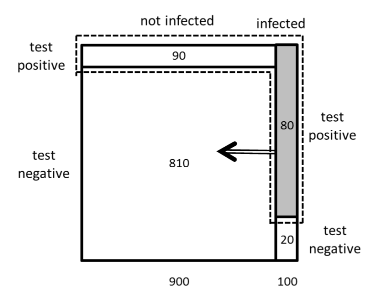
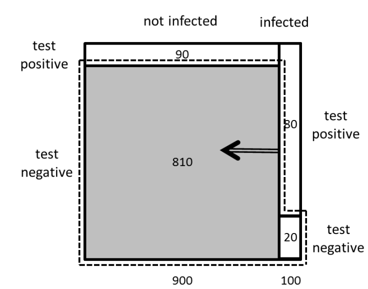
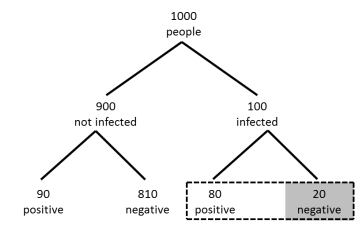
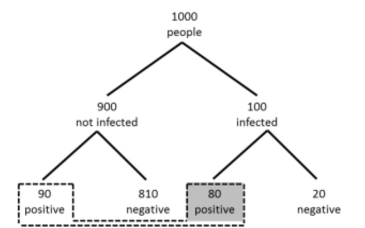
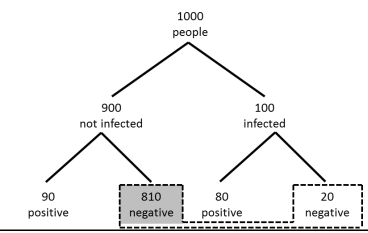
Original test item		
Diagnosis (Study 2) In a preventive medical check-up, 1000 people are tested. The test has the following characteristic: 80% of the infected people and 10% of the uninfected people get a positive test result: How the following proportions change if, in the next routine screening test with 1000 people, the percentage of people infected with the disease is bigger? *Mark the correct solutions.* a. The proportion of people tested negative among the infected will be bigger / smaller / constant. b. The proportion of infected among the people tested positive will be bigger / smaller / constant. c. The proportion of uninfected among the people tested negative will be bigger / smaller / constant.	 	
Explanation of the test item		
Item (a) The proportion corresponding to $P(negative \vert infected)$ is constant if the base rate increases	Item (b) The proportion corresponding to $P(infected \vert positive)$ increases if the base rate increases	Item (c) The proportion corresponding to $P(uninfected \vert negative)$ decreases if the base rate increases
		
 	 	

Figure 6. The Bayesian situation of a diagnosis, items (1) to (3) with numerical solution and visualization by the unit square and the tree.

Moreover, we indicate in the small visualizations the changes of the base rates by arrows. The shape of the tree is not influenced by a change of the base rate, thus there are no arrows in the trees. All items with number (a) in the questionnaire addressed proportions that remain constant when the base rate increases. All items with number (b) addressed proportions that increase with increasing base rates and that correspond to posterior conditional probabilities. All items with number (c) addressed proportions that decreased with increasing base rates and corresspond to posterior conditional probabilities as well.*Results*: Since we rated each item with 1 (correct) or 0 (incorrect) and we had three situations with 3 items each, the maximum value for the questionnaire in this second experiment was 9 points. As in study 1, the Cronbach's Alpha showed an appropriate reliability of the 3 x 3 items and was even slightly better (α = .812). For the sum of scores over the nine items, a t-test yielded the following result: The unit square (*M* = 6.86, *SD* = 2.474) outperformed the tree diagram (*M* = 5.96, *SD* = 2.581), *t*(141) = 2.149, *p* = .033, with an effect size of *d* = .36. Thus, our hypothesis that the unit square is more efficient to support flexible Bayesian reasoning was once more confirmed.

The effect size was slightly better than in study 1, but nevertheless small. But once more, a qualitative analysis of the data showed that this effect was stable regarding different aspects: Figure 7a focused on the different kinds of proportions addressed by the three different items (e.g. "items_a" means smoke_a + diagnosis_a + flowers_a) resulting in a maximum score of 3. Error bars represent the 95% CI of the means. Standard deviations for the tree diagram: items_a (*SD* = 1.187), items_b (*SD* = 1.014), items_c (*SD* = 0.993). Standard deviations for the unit square: items_a (*SD* = 0.977), items_b (*SD* = 0.938), items_c (*SD* = 0.989). The diagram indicates a stable advantage of the unit square for the three different kinds of proportion. Whereas the scores for the tree diagram were almost equal, the benefit of the unit square was the strongest for the proportion (a) that remains constant when the base rate increases. One probable reason might be that proportions that remain constant are more easily to perceive in the unit square than proportions that increase (items_b) or decrease (items_c) with increasing base rates.

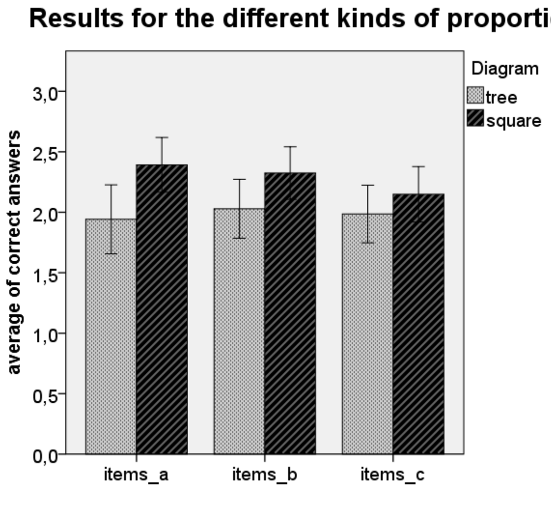


Figure 7a. Results in study 2 for the different kinds of proportion adressed by the different items.

Figure 7b focused on the different Bayesian situations (e.g. "smoke" was obtained by smoke_a + smoke_b + smoke_c, maximum score of 3). Error bars represent the 95% CI of the means. Standard deviations for the tree diagram: smoke (*SD* = 0.909), diagnosis (*SD* = 1.084), flowers (*SD* = 1.244). Standard deviations for the unit square: smoke (*SD* = 0.847), diagnosis (*SD* = 1.054), flowers (*SD* = 1.139). This diagram indicates a stable advantage of the unit square as well, which involves that the benefit of the unit square does not depend on the problem context. However, the scores are a little bit higher in the situation “smoke” for both visualizations which might be due to the familiarity of the students with the problem context.

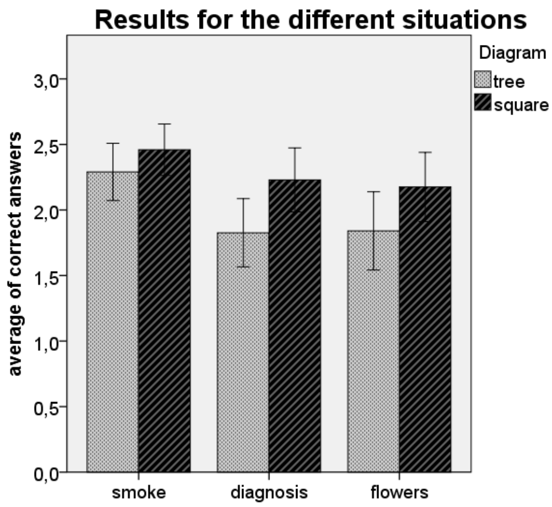


Figure 7b. Results in study 2 for the different Bayesian situations.

7. Discussion

The question of how to promote insight into Bayesian reasoning situations and how to improve for example the comprehension of the influence of the base rate is a crucial issue for mathematics education and has practical consequences for the teaching and learning of statistics. In this article, we focused on flexible Bayesian reasoning which we introduced as the understanding of parameter dependency in Bayesian reasoning situations. We investigated the impact of visualization of statistical information and studied if the tree diagram or the unit square is better suited to boost performance in tasks with variable base rates.

Across two studies, it was consistently found that the unit square outperformed the tree diagram in supporting flexible Bayesian reasoning. Although this difference was small and further research is needed to confirm our findings, we can argue that the measured effect is robust and stable: First, the participants were randomly assigned to the different groups. Second, we measured a similar effect in both experiments for participants which have different backgrounds (students of mathematics education in experiment 1 vs. students of electrical engineering in experiment 2). Third, the effect was found to be stable regarding different problem contexts and different kinds of proportions. Finally, the measured effect was obtained without instruction only by paper-pencil test and we assume that this effect could be enhanced by a short intervention.

The empirical results showing the advantage of the unit square are in line with our theoretical analysis of how base rates are visualized by the unit square and by the tree diagram (see section 4). Since the unit square is a statistical graph, a change of the base rate is directly represented by a changed shape of the unit square. This is not the case for the tree diagram. Thus, if different unit squares were drawn for different base rates, a visual image would be provided for the change of proportions in the data. The tree diagram in contrast represents proportions only in a numerical way. In our studies, where we used test-items with only one visualization for each situation, the participants had to imagine what would change if the base rate increased. Given the higher performance for tasks with the unit square, we can conclude that the unit square encourages a mental representation that is better suited to understand the structure of the data and to imagine the effect of changing base rates.

As mentioned above, we obtained the measured effect without instruction. The participants were not taught how to think about changing base rates or how to imagine changing proportions with the help of the visualizations. We suppose that the beneficial effect of the unit square that was still small in the paper-pencil tests could be enhanced by a short intervention that illustrates how to imagine the change of the visual appearance of the unit square depending of the base rate (see Figure 3). For the tree diagram we would expect no additional improvement by further instructions because the use of the tree diagram was well-known to the participants since the tree diagram is commonly used in mathematics education in Germany. For future research it might be interesting to conduct a training study to prove these assumptions and confirm the advantage of the unit square.

What are the didactical implications of our research findings? First, the construction of the unit square is intuitively understandable. Even without instruction, only by the mean of a simple example illustrating the construction of the unit square, people were able to read beyond the data when estimating the influence of the base rate. As mentioned in the introduction, "Reading beyond the data" is the highest category of graph comprehension formulated by Curcio (1989). For the two lower categories "read the data" and "read between the data" we could show in our previous research that the unit square is equally efficient than the tree diagram (Böcherer-Linder, Eichler & Vogel, 2015). Thus, we conclude that the unit square with natural frequencies is a proper visualization for the teaching and learning of probability and statistics.

Second, based on our theoretical analyses we got empirical evidence that the unit square is slightly more effective than the tree diagram for promoting flexible Bayesian reasoning represented by the understanding of the structure of the data which supports the understanding of the influence of the base rate on conditional probabilities. These are aspects of conceptual understanding in the field of Bayesian reasoning. Thus, our results support the assumption that the unit square seems to be more appropriate than the tree diagram to acquire conceptual knowledge (cf. Hiebert & Lefevre, 1986) of Bayesian reasoning situations.

Third, the unit square can be helpful for the understanding of counterintuitive phenomena in Bayesian reasoning situations (see section 2) since our results show that the unit square is efficient for the estimation of the influence of the base rate. There are some objections that it is not always possible to draw a unit square true to scale because the base rate can become very small (Binder et al., 2015, p. 8). But this is only a limitation of displaying extreme values but not a limitation of developing conceptual knowledge about Bayesian situations when studying the impact of the base rate. Our results show that the unit square supports the understanding of the structure of the data and this can be obtained as well by drawing a unit square approximately. Once the structure of the data is understood with the aid of the unit square, the consideration of extreme base rates is not difficult any more. Our results show that the unit square is efficient when participants perceive the visualization. However, more research is needed concerning the effects for the learning of conditional probabilities when learners are asked to construct the visualization.

First and foremost for further research, it might be interesting to replicate our results with students at school. Since the participants in our studies were undergraduates with no further knowledge about Bayesian reasoning, about the Bayes' rule or about visualizations of statistical information we would expect that the results will not differ when students at secondary school are investigated. Additionally, all our test-items are understandable for students at school, thus, it is possible to directly administer our questionnaire to younger people.

For educational practice and for further research projects, there is a great potential in representing the unit square dynamically with a Computer program which allows the study of the influence of the base rate on conditional probabilities. There are some examples of the unit square with probabilities available (e.g. http://unitsquare.leitideedatenundzufall.de/ [last downloaded 26/11/16]), but we would plead for an application of the unit square with natural frequencies taking into account our research results and the well documented beneficial effect of natural frequencies (e.g. Gigerenzer & Hoffrage, 1995). We would expect that the possibility to explore the influence of the base rate with a dynamic unit square with a computer could further enhance the beneficial effect of the unit square that we proved empirically for the static unit square.

References

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgements. *Acta Psychologica, 44*, 211-233.

Bea, W. (1995). *Stochastisches Denken* [Stochastical Reasoning]. Frankfurt, Peter 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. doi:10.3389/fpsyg.2015.01186.

Böcherer-Linder, K., & Eichler, A. (2015). Vergleich konkurrierender Visualisierungen zum Bereich der bedingten Wahrscheinlichkeiten [Comparison of competing visualizations in the field of conditional probabilities]. In F. Caluori, H. Linneweber-Lammerskitten, & C. Streit (Eds.), *Beiträge zum Mathematikunterricht* (pp. 176 - 179). Münster: WTM-Verlag.

Böcherer-Linder, K. & Eichler, A. (in press). The impact of visualizing nested sets. An empirical study on tree diagrams and unit squares. *Frontiers in Psychology, Cognition.*

Böcherer-Linder, K., Eichler, A., & Vogel, M. (2015). Understanding conditional probability through visualization. In H. Oliveira, A. Henriques, A. P. Canavarro, C. Monteiro, C. Carvalho, J. P. Ponte, R. T. Ferreira, S. Colaço (Eds.), *Proceedings of the International Conference Turning data into knowledge: New opportunities for statistics education* (pp. 14–23). Lisbon, Portugal: Instituto de Educação da Universidade de Lisboa.

Borovcnik, M. (2012). Multiple perspectives on the concept of conditional probability. *AIEM. Avances de Investigación en Educación Matemática, 2*, 5-27.

Brase, G. L. (2009). Pictorial representations in statistical reasoning. *Applied Cognitive Psychology*, *23*(3), 369–381. doi:10.1002/acp.1460.

Brase, G. L. (2014). The power of representation and interpretation: Doubling statistical reasoning performance with icons and frequentist interpretations of ambiguous numbers. *Journal of Cognitive Psychology*, *26*(1), 81–97. doi:10.1080/20445911.2013.861840.

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), 1–73. doi:10.1016/0010-0277(95)00664-8.

Curcio, F. R. (1989). *Developing graph comprehension*. Reston, VA: N.C.T.M.

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

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

Ellis, K. M., Cokely, E. T., Ghazal, S., & Garcia-Retamero, R. (2014). Do people understand their home HIV test results? Risk literacy and information search. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 58* pp. 1323–1327) London: Sage. doi: 10.1177/1541931214581276

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129 – 162). Chichester, UK: Wiley.

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.

Hiebert, J., & Lefevre, P (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and Procedural Knowledge: The case of Mathematics*. (pp. 1 – 27). Hillsdale, NJ: Lawrence Erlbaum.

Hoffrage, U., Gigerenzer, G., Krauss, S., & Martignon, L. (2002). Representation facilitates reasoning: what natural frequencies are and what they are not. *Cognition 84,* 343–352. doi: 10.1016/S0010-0277(02)00050-1.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103* (3), 582 – 591.

Kurz-Milcke, E., Gigerenzer, G., & Martignon, L. (2008). Transparency in risk communication: Graphical and analog tools. *Annals of the New York Academy of Sciences*, *1128*, 18-28.

Mandel, D. R. (2014). The psychology of Bayesian reasoning. *Frontiers in Psychology, Cognition. 5*, 1144. doi: 10.3389/fpsyg.2014.01144.

Navarrete, G., Correia, R., & Froimovitch, D. (2014). Communicating risk in prenatal screening: the consequences of Bayesian misapprehension. *Frontiers in psychology, cognition. 5,* 1272. doi: 10.3389/fpsyg.2014.01272.

Oldford, R. W. (2003). *Probability, problems, and paradoxes pictured by eikosograms.* Occasional paper. Retrieved from www.stats.uwaterloo.ca/~rwoldfor/papers/venn/eikosograms/examples/paper.pdf.

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

Siegrist, M. & Keller, C. (2011). Natural frequencies and Bayesian reasoning: the impact of formal education and problem context. *Journal of Risk Research 14, 1039–1055*. doi: 10.1080/13669877.2011.571786

Spiegelhalter, D., & Gage, J. (2014). What can we learn from real-world communication of risk and uncertainty? In K. Makar (Ed.) *Proceedings of the Ninth International Conference on Teaching Statistics*. Flagstaff, USA: International Statistical Institute. Retrieved from icots.info/icots/9/proceedings/pdfs/ICOTS9_PL2_SPIEGELHALTER.pdf

Sturm, A., & Eichler, A. (2014). Students' beliefs about the benefit of statistical knowledge when perceiving information through daily media. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International*

Conference on Teaching Statistics (ICOTS9), Flagstaff, AR: International Association of Statistics Education. Retrieved from http://icots.info/icots/9/proceedings/pdfs/ICOTS9_7D1_STURM.pdf

Tufte, E. R. (2013). *The visual display of quantitative information* (2nd ed). Cheshire, CT: Graphics Press.

Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 4157* (185), 1124 – 1131.

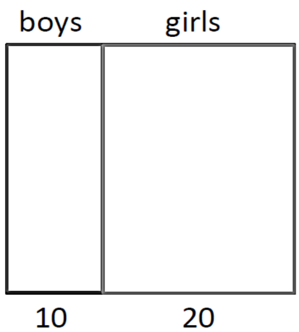
Veaux, R. D. de, Velleman, P. F., & Bock, D. E. (2012). *Intro stats (*3rd ed). Boston MA: Pearson/Addison-Wesley.

Appendix. Translation of the questionnaire with the unit square of study 1 originally in German.

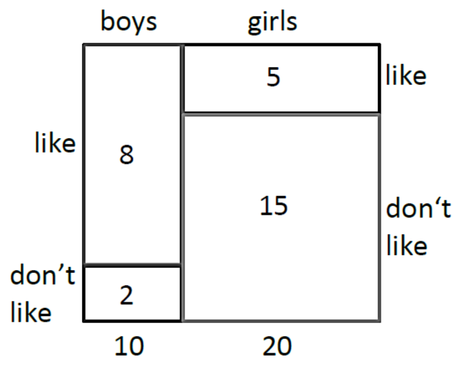
Introductory example: The girls and boys of a school class were interrogated if they like football or not. The results are presented in the following table:

	boys	Girls
like	8	5
don't like	2	15
Sum	10	20

This information can be represented by the unit square, in which the sizes of parts of the square are proportional to the represented numbers. First, the square has to be divided vertically in the proportion of number of boys and number of girls*:*

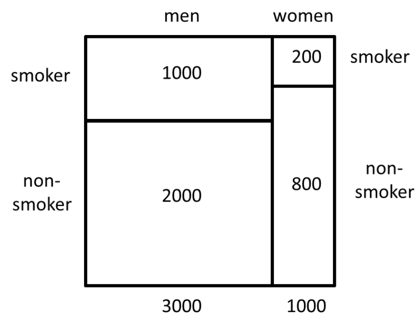


Second, the areas for boys and girls are further divided horizontally in the proportion of numbers of those who like or don't like football.



Please answer to the questions without using a pocket calculator or any other tool.

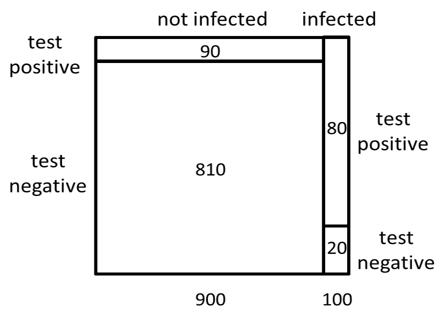
1. Smoke. 4000 students of a university were asked if they smoke or not. It turned out that one-third of the men smoke and one-fifth of the women smoke:



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.

a. The proportion of smokers among the women will be bigger / smaller / constant.
b. The proportion of women among the smokers will be bigger / smaller / constant.
c. The proportion of men who smoke among all students will be bigger / smaller / constant.
d. The proportion of women among the non-smokers will be bigger / smaller / constant.

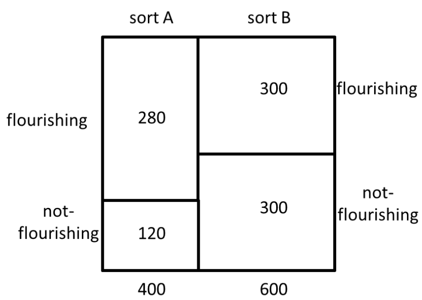
2. Diagnosis. In a preventive medical check-up, 1000 people are tested. The test has the following characteristic: 80% of the infected people and 10% of the uninfected people get a positive test result:



How the following proportions change if, in the next routine screening test with 1000 people, the percentage of people infected with the disease is bigger? *Mark the correct solutions.*

a. The proportion of people tested positive among the infected will be bigger / smaller / constant.
b. The proportion of infected among the people tested positive will be bigger / smaller / constant.
e. The proportion of people tested negative among all tested people will be bigger / smaller / constant.
f. The proportion of infected among the people tested negative will be bigger / smaller / constant.

3. Snowdrops. In autumn, a gardening company has planted 1000 snowdrop bulbs among them were bulbs of the expensive sort A and bulbs of the cheap sort B. In spring, 70% of sort A and 50% of sort B flourish:



How the following proportions change if, in next autumn, the percentage of bulbs of sort A among 1000 planted snowdrop bulbs is bigger? Mark the correct solutions.

a. The proportion of flourishing among bulbs of sort A will be bigger / smaller / constant.
b. The proportion of bulbs of sort A among the flourishing will be bigger / smaller / constant.
c. The proportion of not-flourishing among all bulbs will be bigger / smaller / constant.
d. The proportion of bulbs of sort A among the not-flourishing will be bigger / smaller / c

Referencias a los autores

Katharina Böcherer-Linder[1], University of Education Freiburg (Germany). katharina.boechererlinder@ph-freiburg.de

Andreas Eichler[2], University of Kassel (Germany). eichler@mathematik.uni-kassel.de

Markus Vogel[3], University of Education Heidelberg (Germany). vogel@ph-heidelberg.de

The impact of visualization on flexible Bayesian reasoning

Katharina Böcherer-Linder[1], University of Education Freiburg (Germany).

Andreas Eichler[2], University of Kassel (Germany).

Markus Vogel[3], University of Education Heidelberg (Germany).

Bayesian reasoning is known as notoriously difficult and susceptible to biases, such as e.g. base-rate neglect. Thus, the question of how to promote insight into *Bayesian reasoning situations and how to improve for example the comprehension of the influence of the base rate is a crucial issue for mathematics education and has practical consequences for the teaching and learning of statistics. Research has developed mainly two ideas to support Bayesian reasoning, i.e. to use natural frequencies for the presentation of the statistical information and to visualize the statistical information. However, it is an open question which visualizations are effective for which situations.

In this article we investigate the effects of two visualizations that contain natural frequencies, i.e. the tree diagram and the unit square. We focus on the conceptual understanding of Bayesian reasoning situations and investigate whether the tree diagram or the unit square is more appropriate to support the understanding of the influence of the base rate, which is introduced as being a part of flexible Bayesian reasoning.

As a theoretical background, we firstly briefly outline the results of psychological research concerning the base rate neglect. Subsequently, we show how statistical information and the influence of the base rate on conditional probabilities can be visualized. As a statistical graph, the unit square reflects the influence of the base rate not only in a numerical but also in a geometrical way. Thus, we hypothesize that the unit square is more efficient than the tree diagram to promote the understanding of the influence of the base rate on proportions in the data and therefore to support flexible Bayesian reasoning.

We designed a questionnaire where we asked to estimate the effect of changing base rates for different proportions in the data and conducted two experiments with undergraduate students (N = 148 and N = 143). As hypothesized, the unit square outperformed the tree diagram referring to the understanding of the influence of the base rate in both experiments. Although this difference was small, we show that the measured effect was stable and robust regarding different aspects. Nevertheless, further research is needed to confirm these results, first and foremost it might be interesting to replicate them with students at school. We assume that the beneficial effect of the unit square can be enhanced by a short intervention that illustrates how to imagine the change of the visual appearance to the unit square depending of the base rate.

In the discussion section, we outline the practical consequences for the teaching and learning of statistics. The unit square can be helpful for the understanding of counterintuitive phenomena in Bayesian reasoning situations, since the unit square was shown to be appropriate for the understanding of the influence of the base rate. For educational practice, there is a great potential in representing the unit square dynamically with a computer program which offers the possibility to explore the influence of the base rate in an interactive way.