Thinking in proportions rather than probabilities facilitates Bayesian reasoning

Katharina Loibl
katharina.loibl@ph-freiburg.de
University of Education Freiburg
Kunzenweg 21, 79117 Freiburg, Germany

Timo Leuders
leuders@ph-freiburg.de
University of Education Freiburg
Kunzenweg 21, 79117 Freiburg, Germany

Abstract

Bayesian reasoning tasks require processing data in probabilistic situations to revise risk estimations. Such tasks are difficult when data is presented in terms of single-event probabilities; the multiplicative combination of priors and likelihoods often is disregarded, resulting in erroneous strategies such as prior neglect or averaging heuristics. Proportions (relative frequencies) are computationally equivalent to probabilities. However, proportions are connected to natural mental representations (so-called ratio sense). Mental representations of nested proportions (70% of 20%) allow for a mental operation that corresponds to a multiplicative combination of percentages. In two studies, we focused on the conceptual understanding underlying Bayesian reasoning by utilizing graphical representations without numbers (to avoid calculations with percentages). We showed that verbally framing Bayesian tasks in terms of proportions, as opposed to single-event probabilities, increased correct Bayesian judgment, and reduced averaging heuristics. Thus, we claim, proportions can be regarded as a natural view on normalized Bayesian situations.

Keywords: Bayesian reasoning; averaging heuristics; biased strategies; mental representations; nested proportions

Introduction

Bayesian reasoning is a type of inference that revises the subjective probability of hypotheses *H* on the world by considering new data *D* and its likelihood. When and why humans are capable of such inferences has been investigated extensively in situations with quantifiable data. One of the most prominent situations asks for the probability of a disease (*H*) after a positive test result (*D*) using information on the prevalence of the disease as well as the test's sensitivity and specificity (Eddy, 1982; Gigerenzer & Hoffrage, 1995). Similarly, teachers may judge the probability of a misconception in a student (*H*) after observing an erroneous solution (*D*) using information on the prevalence of the misconception at that grade level as well as information on the task (Leuders & Loibl, 2020).

While most research investigates Bayesian reasoning with precise numbers, in many situations, one may only have access to approximate estimations of these values and make a judgment by qualitative reasoning: e.g., "a large part of the small part of patients with the disease is tested positive – which proportion of the positive tested patients actually has the disease?" The situation can also be verbally framed in probabilities for a single case: "There is a small probability that a patient has this disease and there is a high probability that a patient with the disease is tested positive – what is the probability that a positively tested patient has the disease?"

In both framings the step of multiplicatively combining the two probabilities (e.g., "70% of 20%", or, when no numerical values are given: "a small proportion of a large proportion") is crucial for producing a correct estimation. However, without support, few people correctly combine two probabilities multiplicatively in conjunctive situations (Gigerenzer & Hoffrage, 1995; Cosmides & Tooby, 1996; Juslin et al. 2015).

In research on Bayesian reasoning, many elements of the information presentation have been varied systematically in order to disentangle the relevant cognitive processes (cf. McDowell & Jacobs, 2017). So far, the focus has been predominantly on the difference between probabilities and natural frequencies. Solution rates increase drastically when the data is presented as natural frequencies (e.g., "140 of 1000 patients have a disease and of those 80 are tested positive"). However, with natural frequencies the base rate (prevalence) and the hit rate (test specificity) are already contained in the given information ("10 out of 140 of 1000"), eliminating the necessity of the step of combining them. We do not further elaborate on the distinction between absolute and natural frequencies since our research focus on the comparison of probabilities and proportions (relative frequencies), both requiring the combination step.

In our research, we directly compare a framing of single-event probabilities with a proportion framing. In the proportion framing the verbal description ("part of" wording, multiple cases indicated by plural) can activate mental representations of part-whole ratios and, more specifically, the representation of "nested proportions" (proportions of proportions) which allow for a mental operation that corresponds to a multiplicative combination of the base rate and hit rate. So far, only few studies focused on the impact of proportions (usually termed relative frequencies). This can be ascribed to the fact that it is hard to distinguish cognitive representations and processes induced by probabilities and proportions, when they are presented numerically: A probability of 40% may be immediately translated into the proportion 40 out of 100.

To address this gap, we (1) analyze the literature that compares probability and proportion and (2) develop a study design (and implement it in two studies with different contexts) that allows for a clearer theoretical and empirical distinction of the thinking connected to probability and proportion.

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

Bayesian reasoning with proportions

Generally, one may conclude that research has shown that individuals' capacity for Bayesian reasoning, i.e., for processing information on the prior and conditional values of data, depends on the conceptual framing. A conceptual framing of a Bayesian reasoning situation comprises all elements of a given situation that induce the activation and use of certain concepts, in our case probabilities, proportions (relative frequencies), or absolute frequencies (cf. Table 1).

Table 1: Cues within probability, proportion (relative frequency), and absolute frequency framing.

Probability framing	Proportion (relative frequency) framing	Absolute frequency framing
Single event "a 20% chance" "a 0.2 probability"	Multiple events/ sample "20% of all cases" "1/5 of all cases"	Multiple events/ sample "200 out of 1000"

The most frequently studied variation of the conceptual framing is the comparison of Bayesian reasoning with either probabilities or natural frequencies (cf. McDowell & Jacobs, 2017). Proportions are much less explicitly addressed in research on Bayesian reasoning. As with probabilities, data is presented as percentages or fractions and is also normalized and, thus, requires a multiplicative combination of base rates and hit rates. In contrast to single-event probabilities, proportions refer to multiple cases from an abstract sample (i.e., "80% of all ..."). These differences between probabilities and proportions may result in different mental representations.

Mental representations and operations during Bayesian reasoning

Bayesian situations can be regarded as (mathematical) word problems (Johnson & Tubau, 2015; McDowell & Jacobs, 2017; de Corte et al., 2000). A crucial factor in solving word problems is the construction of a mental representation of the situation and the mental operations that lead to a solution. When the mental representations bear structural analogies to the world, individuals are able to mentally read off information or draw inferences via mental manipulation (Johnson-Laird, 1983, Gentner & Stevens, 1983; Vosniadou, 2002, for mathematics see Prediger, 2008; Fischbein, 1989; vom Hofe & Blum, 2016; Thevenot, 2010). For the case of Bayesian reasoning with proportions or probabilities the relevant steps are the following (similar to the "cognitive algorithms" by Gigerenzer & Hoffrage, 1995, who however focus on Bayesian reasoning with natural frequencies vs. probabilities).

Step 1: Mental representation: Structure of the situation and quantitative values The mental representation of the structure of the situation requires an understanding of its relevant constituents and their relationships. In Bayesian reasoning, this means recognizing a specific subset configuration (Tversky & Kahneman, 1983; Gigerenzer & Hoffrage, 1995; Girotto & Gonzales, 2001), sometimes referred to as nested set structure (Sloman et al., 2003). When presenting a Bayesian situation with proportions, the subset structure is made salient via the part-of-a-whole relationship that is inherent in the proportion concept through phrases like "of all" and "of these". In particular, the specific subset structure of Bayesian reasoning can be characterized as "nested proportions" or "part-of-part relation". However, due to the normalization – which is inherent in the proportion concept – the sizes of the part and the whole are not explicit (conversely, a presentation of natural frequencies explicitly shows the sizes). A probability framing activates different mental representations: The single event formulation ("probability/chance that a person") and the sequential process of a random experiment do not directly suggest the subset structure.

In addition to the structure, the (relative) size of the values need to be represented mentally. There is ample evidence that humans (and even primates) are capable of discerning and discriminating continuous magnitudes and ratios (proportions) in various formats and modalities (Bonn & Cantlon, 2017; Jacob et al., 2012; Park et al., 2020). Therefore, proportions can be assumed to have a non-symbolic mental representation. The mental representation of the size of probabilities, however, is less clear (Juslin et al., 2015).

Step 2: Mental operation: Combining normalized information Our tenet is, that mental representations of proportions as part-whole ratios can support mental operations to combine the given information: When the base rate and hit rate are mentally represented as part-whole ratios, one can see the part in the base rate (e.g., patients with disease, students with misconception) as the whole in the hit rate (positively tested patients among those with disease, students with error among those with misconception). Such a mental representation of a ratio within a ratio (i.e., nested proportions) allows for drawing inferences on the magnitude; for example, a small part of a large part is similar to a large part of a small part. This combination of base rate and hit rate is adequate for Bayesian reasoning in normalized situations. It is equivalent to a formal multiplication of base rate and hit rate when they are given by numbers. In contrast, such a multiplicative interaction is not intuitive when values are framed as single-event probabilities ("a chance of 95% after a chance of 20%"). When individuals do not construct a mental representation of a situation of nested proportions, they may resort to using heuristics, such as determining the average (instead of the product) of the two values, to combine the two probabilities.

Step 3: Mental operation: Determining requested ratio Finally, Bayesian reasoning requires two mental operations to determine the requested ratio. First, the two nested proportions, the results from the previous step, must be joined to a new whole. Then, the ratio of one part to the new whole must be determined. Both operations can be considered as mental operations within the mental representation of proportions based on the ratio sense described above. Recognizing a ratio independent of the absolute size of the constituents defines the human capacity to identify part-whole ratios, as found in many studies (e.g., Matthews & Ellis, 2018). The mental steps as described above are necessary for Bayesian reasoning in normalized situations with a proportion or probability framing (cf. Table 2).

Table 2: Mental steps in Bayesian reasoning with proportion or probability framing.

Step	Proportion framing	Probability framing
1	Given: 20% of all cases are *H* 70% of *H* cases show data *D* 10% of non-*H* cases show data D Inferred: Non-*H* part of all cases: 100% – 20% = 80% Required: Which proportion of cases with *D* are also *H* cases?	Given: 20% chance of *H* 70% chance of *D* given *H* 10% chance of *D* given non-*H* Inferred: Chance of non-*H*: 100% – 20% = 80% Required: What is the chance of *H* for a single case with *D*?
2	70% with *D* of 20% *H* → 14% 10% with *D* of 80% non-*H* → 8%	20% chance of *H* and then 70% chance of *D* → 14% 80% chance of non-*H* and then 10% chance of *D* → 8%
3	All cases with *D*: 14% + 8% = 22% Proportion of *H* cases of all cases with *D*: 14% of 22% → 63%	Chance for *D*: 14% + 8% = 22% Chance of *H* for a single case with *D*: 14% / 22% → 63%

Research on different conceptual framing

Research on Bayesian reasoning has examined the effects of different conceptual framings (mostly probability and natural frequency, but also proportion) and different sampling structures (normalized or naturally sampled) (cf. Table 3).

Table 3: Overview of different conceptual framings and sampling structures. Our studies compare ④ vs. ⑤.

		Conceptual framing		
		Probability (single case)	Proportion (multiple cases, abstract sample)	Frequency (multiple cases, concrete sample)
Sampling structure	Normalized	①	②	③
	Non-normalized	④	⑤	⑥

The most frequently studied variation is the comparison of Bayesian reasoning with either normalized probabilities (①) or natural frequencies (⑥) (cf. McDowell & Jacobs, 2017). However, the effects of the different conceptual framings (single-event probability vs. multiple-case absolute frequencies) and the reduction of computational complexity often remain intertwined in this design (Brase & Barbey, 2006).

Macchi (2000) describes two conditions that contribute to the comparison of probability and proportion: In a so-called "partitive condition", the subset structure was made explicit and also the numerical values were given with proportions of multiple cases (②; "360 out of 1000, 75% of these" referred to as probability by the author). In a so-called "non-partitive" condition, the subset structure was implicit, and the numerical values were given as single-case probabilities (①). Thus, the better results in the partitive condition may – as Macchi (2000) argues – be attributed to the supported recognition of the subset structure (cf. mental step 1). It may, however, also be caused by the facilitated combination of normalized information via nested proportions (cf. mental step 2).

Gigerenzer and Hoffrage (1995, study 2) also compared probabilities (①) to proportions (②, in the form of relative frequencies) represented as percentages. The verbal descriptions systematically differed with regard to probabilities versus relative frequencies. However, the questions were always posed in terms of single-event probabilities. Using this design, they found no differences between probability and proportion with respect to the numerical solution procedure and solution rates. As previously mentioned, the symbolic arithmetic procedure is the same for both proportions and probabilities. Moreover, the question format may have triggered all participants to switch to thinking in probabilities.

Indeed, Weber et al. (2018), using a design adopted from Gigerenzer and Hoffrage (1995), showed that participants very often switch between the concepts of probability/proportion and absolute frequency. The concepts and procedures actually used by the participants predicted the rates of successful Bayesian reasoning much more accurately than did the conceptual framing provided in the task.

These findings strongly suggest that a modified research approach is needed to better separate the effects of the computational procedure from the conceptual reasoning process. We therefore attempted to minimize internal switches between probabilities and proportions by implementing consistent conceptual framings throughout the task (representation format, verbal description, question, and answer format) and by avoiding symbolic representations (e.g., percentages), since these can trigger either implicit transitions between probabilities and proportions ("percent of") or procedural knowledge ("multiply percentages").

Strategies and heuristics employed during the multiplicative step of Bayesian reasoning

In normalized Bayesian situations, the prior values (i.e., prior probabilities or base rates) and conditional values (i.e., likelihoods or hit rates) have to be combined multiplicatively to derive the posterior values. However, such a multiplicative interaction is not intuitive and is therefore cognitively demanding (Sundh, 2019). Unsurprisingly, research shows that humans often fail to apply the Bayes rule correctly, even when strongly supported (Weber et al., 2018). Research revealed multiple strategies that deviate from standard Bayesian updating, such as neglecting part of the information (Gigerenzer & Hoffrage, 1995, see also Bruckmaier et al., 2019; Leuders & Loibl, 2020, for overviews on these incorrect strategies). Among these are various evidence-only strategies (EOS; also known as base rate neglect) or prior-only strategies (POS). Moreover, when considering all pieces of information, people tend to combine the information additively as opposed to multiplicatively. In an early study, Shanteau (1975) showed that, when updating probability estimations based on non-informative evidence, the probability updates of the participants suggest averaging instead of multiplying strategies (cf. Loibl & Leuders, 2020: averaging-priors-and-evidence strategy, APES). Similar additive strategies have been identified for joint probabilities (Sundh, 2019; Juslin et al., 2015) and Bayesian reasoning (Cohen & Staub, 2015; Juslin et al., 2009; Lopes, 1985; Macchi, 2000; Shanteau, 1975). Such erroneous strategies can also be seen in mathematics education: When solving word problems, students often regress to additive strategies when unable to construct a situation model with a multiplicative structure (Verschaffel et al., 2020).

The choice of strategy for combining prior and conditional values is influenced by the representation of the situation. Proportions support mental representations of nested proportions, thereby suggesting a combination that is equivalent to multiplicative reasoning. On the other hand, multiplicatively combining probabilities in consecutive events is not obviously supported by a mental representation. We consider this discrepancy between proportions and probabilities to be an explanation for the different tendencies to apply either the multiplicative Bayesian update strategy (BUS) or an additive strategy such as APES. However, the effects of framing the situation in terms of single-event probabilities or proportions on the use of additive and multiplicative strategies have rarely been investigated.

Research question

Our two studies attempt to complement the existing research by a systematic consideration of Bayesian reasoning within a *proportion* framing. In both studies, we (1) implement strictly parallel proportion and probability conditions, (2) provide information that equally requires multiplicative reasoning in both conditions (i.e., normalized samples), and (3) rely on non-symbolic presentations of quantitative values in order to exclude solutions based on (often superficial) procedural calculations.

We assume that framing Bayesian tasks in terms of proportions as opposed to single-event probabilities improves Bayesian estimations of posteriors because – as described above – it activates mental representations of nested proportions which allow for mental operations that correspond to a multiplicative combination of given quantities. Conversely, we assume that framing Bayesian tasks in terms of single-event probabilities as opposed to proportions is more likely to result in inadequate additive combinations of prior and conditional values.

More specifically, we hypothesize that different conceptual framings result in different distributions of updating strategies, namely a predominance of the Bayesian update strategy (BUS) within a proportion framing and a predominance of the averaging-priors-and-evidence strategy (APES) within a single-event probability framing.

Methods

Participants

The participants in study 1 were 37 students enrolled in a mathematics teacher education program from two parallel, identical courses on statistics. In this course, they had not yet worked on probability theory at the time of the study. Each course was randomly assigned to one of the two conditions, either proportion framing ($N = 21$) or probability framing ($N = 16$). The groups did not differ with regard to age ($M = 22.4$, $SD = 1.3$), semester ($M = 5.6$, $SD = 1.5$), nor in their self-reported competence in proportions ($M = 2.7$, $SD = 0.7$) or probability ($M = 2.5$, $SD = 0.7$) on a scale from 1 to 4.

The participants in study 2 were 61 11th graders with a major in mathematics from a vocational high school. They had not worked on probability theory prior to the study. Individual students were randomly assigned to one of the two conditions, either proportion framing ($N = 31$) or probability framing ($N = 30$). The groups did not differ with regard to age ($M = 17.4$, $SD = 0.6$) nor in their self-reported competence in proportions ($M = 2.4$, $SD = 0.8$). However, students in the probability condition reported a slightly higher competence in probability ($M = 2.7$, $SD = 0.8$ vs. $M = 2.2$, $SD = 0.6$) on a scale from 1 to 4.

In both studies, all participants (and for the underaged students their legal guardians) provided written consent to take part in the study.

Design

We investigated Bayesian reasoning in a non-symbolic setting. That is, all relevant pieces of information (prior values, conditional values, and posteriors) were represented graphically and qualitatively with bar charts on a computer screen and without symbols (neither numbers, variables, nor formulas, cf. Fig. 1). The representation enables the participants to (mentally and practically) manipulate the bars in a way that corresponds to the combination step (either correct multiplication or biased addition). This is done without already displaying the result of the combination step, which is the case in most graphical representations, e.g., a unit square that displays the combined sizes (Eichler et al., 2020; Khan et al., 2015). And, most importantly, the bars can be verbally described using both framings: probabilities and proportions and, thus, be described identically for both conditions.

The context of the presented Bayesian tasks in study 1 (participants: teacher students) described a population of learners having problems with decimals and the participants were asked to decide between one of two possible (and very common, e.g., Moloney & Stacey, 1997) misconceptions – the shorter-is-larger misconception (SL) and the whole-number misconception (WN) – after data in the form of an erroneous student response to a specific test question is known. In study 2 (participants: high school students), we used the medical context of testing for the flu, as often applied in Bayesian reasoning research.

In order to systematically vary the conceptual framing, the wording (presented both verbally and through onscreen labels) was adapted across conditions (e.g., study 1: “proportion of students with SL” vs. “probability of a student for SL” and “proportion of students with error of students with SL” vs. “probability of a student for an error when SL”; study 2: “proportion of patients with the flu” vs. “probability of a patient for the flu” and “proportion of patients with the flu of the positive tested patients” vs. “probability of a patient for the flu when tested positive”).

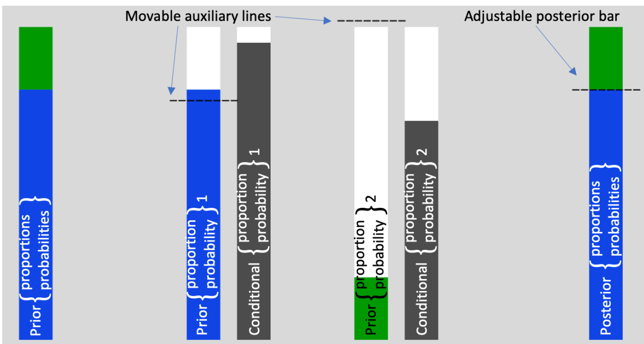

Figure 1: Structure of the graphical updating environment (labels for participants different, see example in text).

Prior to data collection, participants received an instruction about the tasks and the meanings of the bar charts. Afterwards, participants had to judge eight cases in study 1 and six cases in study 2. Each case was made up of an unknown student or patient, a test question or a medical test, and the information that the given student’s response was incorrect or the patient was tested positive. Each case presented the prior values (green and blue bars to the left in Fig. 1). These values were presented again in two separate bars, with the likelihoods/hit rates of an error or a positive flu test given in two adjoining bars (gray and white bars in Fig. 1). Participants were prompted to combine the prior and conditional information and to use auxiliary marker lines (dashed horizontal lines) to support their thinking. Finally, they were asked to estimate the posteriors by adjusting the posterior bar (the far-right bar in Fig. 1) for the eight (or in study 2: six) successively presented cases. The wording of the instructions and the graphical display differed only with respect to probability and proportion language. The graphical display equally allowed for participants in both framings to represent their strategy of combining information. After the final case, participants described their strategy in a stimulated recall.

The participants’ estimation of the posterior values for the eight cases enabled us to distinguish between their updating strategies. To this end, the prior and conditional values for the cases were selected in a way that the posteriors resulting from the possible updating strategies differed maximally – especially with regard to BUS and APES (see next section).

Classification of updating strategy

To classify the participants with respect to their updating strategies (Bayesian update strategy BUS, averaging-priors-and-evidence strategy APES, evidence-only strategy EOS, or priors-only strategy POS), we inspected their posterior estimations (i.e., their selected positions in the posterior bars) in the eight (or in study 2: six) estimation tasks and used the distance from the exact values for classification.

As participants were not expected to hit the exact posterior position, an estimated posterior value (e.g., 60%) cannot be attributed with certainty to a particular updating strategy. Instead, an estimated posterior value contributes, at different levels of strength, to the plausibility of each of the four strategies. Therefore, we followed a naïve Bayesian classification procedure (Duda et al., 2012); we modeled the likelihood of a subject’s judgment J (i.e., estimation of posterior) under the condition that he or she applied a specific updating strategy (e.g., BUS) using a Gaussian distribution. The likelihood of an estimation value J_{ij} of subject j in task i decreases exponentially as the distance to the mathematically exact judgment increases.

The evidence from eight (or in study 2: six) estimation tasks was used to update an (unknown) prior probability for a strategy according to the product $p^{post}(\mathrm{BUS}) = p^{prior}(\mathrm{BUS}) \cdot \prod_i p_i(J_{ij}|\mathrm{BUS})$. The multiple ratios (read as pairwise ratios) express the change in probability for the classification induced by the accumulated evidence and are known as Bayes factors (BF):

$$\prod_i p_i(J_{ij}|\mathrm{BUS}) : \prod_i p_i(J_{ij}|\mathrm{APES}) : \prod_i p_i(J_{ij}|\mathrm{POS}) : \prod_i p_i(J_{ij}|\mathrm{EOS})$$

e.g. $\mathrm{BF}_{\mathrm{BUS:APES}}$ $\mathrm{BF}_{\mathrm{APES:EOS}}$

For a classification of the updating strategy of each subject j, we inspected the Bayes factor of the dominant strategy to the subsequent one, e.g., $\mathrm{BF}_{\mathrm{BUS:APES}}$. The resulting certainty of classification equals the posterior ratio, assuming equal priors.

Differences in the distribution of the updating strategies BUS and APES between the experimental conditions were analyzed using a Bayesian contingency table test with an independent multinomial model (Jamil et al., 2017) via JASP software.

Results

Most participants in study 1 could be classified with high certainty, mostly with extreme evidence, $\mathrm{BF}_{1:2} > 100$. $\mathrm{BF}_{1:2}$ indicates the increase of the likelihood of one classification over the other (the ratio of the dominant classification to the subsequent one). The classification for five participants was uncertain ($\mathrm{BF}_{1:2} \leq 3$). Table 4 lists the classification of the participants to BUS, APES, or one of the other strategies (POS, EOS) after eight cases. When comparing the distributions of BUS and APES across conditions, a Bayesian contingency table test revealed moderate evidence ($\mathrm{BF}_{10} = 6.1$) for our hypotheses: Participants in the proportion framing condition tended to apply more BUS, participants in the probability framing condition tended to apply more APES.

Most participants in study 2 could be classified with high certainty. The classification for 16 participants was uncertain

($BF_{1:2} \leq 3$). Table 5 lists the classification of the participants to BUS, APES, or one of the other strategies (POS, EOS) after six cases. In contrast to study 1, many participants failed to combine the information and recurred to single-information strategies (POS or EOS). When comparing the distributions of BUS and APES across conditions, a Bayesian contingency table test revealed moderate evidence ($BF_{10} = 5.4$) for our hypotheses: Participants in the proportion framing condition tended to apply more BUS, participants in the probability framing condition tended to apply more APES.

Table 4: Number of participants per condition classified to the strategies with high certainty ($BF_{1:2} > 3$) in study 1.

	BUS	APES	others	$BF_{1:2} \leq 3$
Proportion	13	3	2	3
Probability	5	8	1	2

Table 5: Number of participants per condition classified to the strategies with high certainty ($BF_{1:2} > 3$) in study 2.

	BUS	APES	others	$BF_{1:2} \leq 3$
Proportion	8	2	13	8
Probability	3	7	11	9

Discussion

The main assumption of our studies was that the cognitive processes constituting Bayesian updating in non-symbolic situations can rely on mentally representing and processing proportions as parts of parts. We assumed that this mental representation and operation is at work when the relevant information (prior and conditional values) is presented in a proportion framing. In contrast, a probability framing may not activate this parts-of-parts thinking to a similar extent. As hypothesized, our results show that in both studies participants in the proportion framing condition tended to apply the valid Bayesian update strategy (BUS), whereas participants in the probability framing condition tended to apply a biased strategy. More specifically, participants in the probability framing condition often combined the prior and conditional probabilities additively (by averaging priors and evidence information, APES) instead of multiplicatively. Therefore, we regard our results as supportive for the assumption that parts-of-parts thinking facilitates correct multiplicative Bayesian updating. Additive strategies (similar to APES) have also been detected by Cohen and Staub (2015), Juslin et al. (2009), Lopes (1985), and Shanteau (1975), although these studies could not ascribe the findings systematically to a probability vs. proportion framing.

The substantial increase of BUS in the proportion framing condition in our studies is especially noteworthy when compared to prior research on fostering Bayesian reasoning. Approaches that use natural frequencies instead of probabilities (e.g., Gigerenzer & Hoffrage, 1995; Hill & Brase, 2012), not only highlight the parts-of-parts structure (step 1) but also present the conditional values as joint frequencies (e.g., 2 of the 10 students with SL solve this task correctly). Thus, the priors (e.g., 10 of 100 students) are already contained in the joint frequencies, eliminating the necessity of the crucial step 2 (determining the parts of parts). Similarly, graphical representations such as the unit square (Eichler et al., 2020; Khan et al., 2015) explicitly present the parts-of-parts structure (step 1), but also the magnitude of the resulting parts of parts (step 2). In contrast, our proportion framing still requires all three steps. Thus, we could show an improvement in Bayesian reasoning without reducing the complexity.

Our investigation was – different from most approaches – conducted in a non-symbolic setting by providing the relevant information in the form of bar charts instead of numbers. Previous studies also included graphical representations to visualize the structure of the situation (e.g., Eichler et al., 2020; Kahn et al., 2015), but these studies usually also provided numerical information and do not rely solely on graphical representations. Therefore, in these studies the expected valid strategy amounted to applying the Bayes rule by calculation. As known from research on word problems, providing numbers can lead to superficial calculations without constructing a situation model (for a review see Verschaffel et al., 2020). Indeed, other studies (for Bayesian reasoning, Gigerenzer & Hoffrage, 1995; for conjunctive probabilities, Juslin et al., 2009) found incorrect calculations that corresponded to ad-hoc strategies in combining the provided numbers. A non-symbolic approach can help to reduce such interferences from procedural calculations. In addition, non-symbolic Bayesian reasoning also seems ecologically valid in many real-life contexts, such as teachers' judgments of students' misconceptions, in which the information is rather not represented symbolically, but only by qualitative estimations, and thus, the process of Bayesian reasoning also relies on processing such information qualitatively (Leuders & Loibl, 2022).

In our studies, we investigated the cognitive processes involved in posterior estimation. In situations where only a decision for the most likely hypothesis (e.g., which misconception or disease) is required, the steps in our model can be simplified. Loibl and Leuders (2020) have shown in a simulation that APES leads to correct decisions in most cases. Whether this phenomenon is also shown empirically still needs to be investigated.

To conclude, our studies show that it is promising to systematically investigate the effect of recurring to mental representations of proportions in Bayesian reasoning. With a focus on the parts-of-parts model, our studies investigated step 2 (determining the parts of parts) of the cognitive model that is central to Bayesian updating. Further research should also investigate the other steps within the cognitive model. This endeavor requires a fundamental analysis of the interaction between representations and cognition as already put forward by Gigerenzer and Hoffrage (1995). In this regard, our studies suggest that investigating non-symbolic settings is a promising strategy to enlighten the cognitive processes at work in Bayesian reasoning.

References

Bonn, C. D., & Cantlon, J. F. (2017). Spontaneous, modality-general abstraction of a ratio scale. *Cognition, 169,* 36-45.

Brase, G. L., & Barbey, A. K. (2006). Mental representations of statistical information. *Advances in psychology research*, *41*, 91-113.

Bruckmaier, G., Binder, K., Krauss, S. & Kufner H.-M. (2019). An eye-tracking study of statistical reasoning with tree diagrams and 2 × 2 tables. *Frontiers Psychology, 10*, 632.

Cohen, A. L., & Staub, A. (2015). Within-subject consistency and between-subject variability in Bayesian reasoning strategies. *Cognitive psychology*, *81*, 26-47.

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.

de Corte, E., Greer, B., & Verschaffel, L. (Eds.). (2000). *Making sense of word problems.* CRC Press.

Duda, R. O., Hart, P. E., & Stork, D. G. (2012). *Pattern classification.* John Wiley & Sons.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: problems and opportunities, In D. Kahneman, P. Slovic, and A. Tversky (Eds.)*, Judgment under Uncertainty: Heuristics and Biases* (p. 249-267). New York, NY: Cambridge University Press.

Eichler, A., Böcherer-Linder, K., & Vogel, M. (2020). Different visualizations cause different strategies when dealing with Bayesian situations. *Frontiers in Psychology, 11,* 1897.

Fischbein, E. (1989). Tacit models and mathematical reasoning. *For the learning of mathematics*, *9*(2), 9-14.

Gentner, D. & Stevens, A.L. (1983). *Mental models*. Hillsdale, NJ: Lawrence Erlbaum.

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

Girotto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: a matter of information structure and question form. *Cognition, 78,* 247-276.

Hill, W., & Brase, G. (2012). When and for whom do frequencies facilitate performance? On the role of numerical literacy. *The Quarterly Journal of Experimental Psychology, 65*, 2343-2368.

Jacob, S. N., Vallentin, D., & Nieder, A. (2012). Relating magnitudes: the brain's code for proportions. *Trends in cognitive sciences, 16*(3), 157-166.

Jamil, T., Ly, A., Morey, R., Love, J., Marsman, M., & Wagenmakers, E. J. (2017). Default "Gunel and Dickey" Bayes factors for contingency tables. *Behavior Research Methods, 49,* 638-652.

Johnson, E.D., & Tubau, E. (2015). Comprehension and computation in Bayesian problem solving. *Frontiers in Psychology, 6,* 938.

Johnson-Laird, P. N. (1983). Mental models. *Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.

Juslin, P., Lindskog, M., & Mayerhofer, B. (2015). Is there something special with probabilities? Insight vs. computational ability in multiple risk combination. *Cognition, 136*, 282-303.

Juslin, P., Nilsson, H., & Winman, A. (2009). Probability theory, not the very guide of life. *Psychological Review, 116*(4), 856-874.

Khan, A., Breslav, S., Glueck, M., & Hornbæk, K. (2015). Benefits of visualization in the mammography problem. *International Journal of Human-Computer Studies, 83,* 94-113.

Leuders, T., & Loibl, K. (2020). Processing probability information in nonnumerical settings–teachers' Bayesian and non-Bayesian strategies during diagnostic judgment. *Frontiers in Psychology*, *11*, 678.

Loibl, K., & Leuders, L. (2020). "Take the middle" – Averaging prior and evidence as effective heuristic in Bayesian reasoning. In S. Denison., M. Mack, Y. Xu, & B.C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 1764-1770). Cognitive Science Society.

Lopes, L. L. (1985). Averaging rules and adjustment processes in Bayesian inference. *Bulletin of the Psychonomic Society, 23*, 509-512.

Macchi, L. (2000). Partitive formulation of information in probabilistic problems: Beyond heuristics and frequency format explanations. *Organizational behavior and human decision processes, 82*(2), 217-236.

Matthews, P. G., & Ellis, A. B. (2018). Natural alternatives to natural number: The case of ratio. *Journal of numerical cognition*, *4*(1), 19-58.

McDowell, M., & Jacobs, P. (2017). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin*, *143*(12), 1273-1312.

Moloney, K., & Stacey, K. (1997). Changes with age in students' conceptions of decimal notation. *Mathematics Education Research Journal*, *9*(1), 25-38.

Park, Y., Viegut, A. A., & Matthews, P. G. (2020). More than the sum of its parts: Exploring the development of ratio magnitude versus simple magnitude perception. *Developmental Science, 24*(3), e13043.

Prediger, S. (2008). The relevance of didactic categories for analysing obstacles in conceptual change: Revisiting the case of multiplication of fractions. *Learning and instruction*, *18*(1), 3-17.

Shanteau, J. (1975). Averaging versus multiplying combination rules of inference judgement. *Acta Psychologica, 39*, 83-89.

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes*, *91*(2), 296-309.

Sundh, J. (2019). The cognitive basis of joint probability judgments. Processes, ecology, and adaption. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Social Sciences, 166*. Uppsala: Acta Universitatis Upsaliensis.

Thevenot, C. (2010). Arithmetic word problem solving: evidence for the construction of a mental model. *Acta Psychologica, 133*, 90-95.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293-315.

Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: a survey. *ZDM Mathematics Education, 52*(1), 1-16.

Vom Hofe, R., & Blum, W. (2016). "Grundvorstellungen" as a category of subject-matter didactics. *Journal für Mathematik-Didaktik, 37*(1), 225-254.

Vosniadou, S. (2002). Mental models in conceptual development. In L. Magnani & N. J. Nersessian (Hrsg.), Model-based reasoning. *Science, Technology, Values* (pp. 353–368). Boston: Springer.

Weber, P., Binder, K., & Krauss, S. (2018). Why can only 24% solve Bayesian reasoning problems in natural frequencies: Frequency phobia in spite of probability blindness. *Frontiers in Psychology, 9*, 1833.