Educational Studies in Mathematics
https://doi.org/10.1007/s10649-024-10372-y

Students' performance and typical errors in filling empty probabilistic visualizations with probabilities or frequencies

Michael Rößner[1] · Karin Binder[1] · Corbinian Geier[2] · Stefan Krauss[3]

Accepted: 11 November 2024
© The Author(s) 2025

Abstract

It has been established that, in Bayesian tasks, performance and typical errors in *reading* information from filled visualizations depend both on the type of the provided visualization and information format. However, apart from reading visualizations, students should also be able to create visualizations on their own and successfully use them as heuristic tools in modeling tasks. In this paper, we first want to broaden the view on Bayesian reasoning to probabilistic tasks with two binary events in general and embed the whole process of solving these tasks using probabilistic visualizations in a modified modeling framework. Thereby, it becomes apparent that most of the steps remained untouched by existing research. Second, in the present empirical study, we focused on one part of the largely unexplored creation process and examined *entering* statistical information into empty visualizations as heuristic tools. $N = 172$ participants had to enter conditional and joint probabilities or the corresponding frequencies into empty visualizations in a paper-and-pencil test. We analyze (a) students' performance when entering information in visualizations and (b) typical errors, both dependent on the *information format* (probabilities vs. natural frequencies), which *empty visualization structure* (2×2 table, double tree, net diagram) was provided, and *type of information* (conditional vs. joint information). The well-known positive effect of natural frequencies on participants' performance was evident when entering conditional information into 2×2 tables and net diagrams. However, with respect to joint information, no superior effect of frequencies was observed. Furthermore, the theoretical implementation of our research in a modeling cycle allows us to identify desiderata for future research.

Keywords Filling empty visualizations · Net diagram · Double tree · 2×2 table · Frequency effect · Typical errors

✉ Michael Rößner
roessner@math.lmu.de

1 Ludwig-Maximilians-Universität in Munich, Munich, Germany

2 Descartes-Gymnasium, Neuburg an der Donau, Germany

3 University of Regensburg, Regensburg, Germany

Published online: 06 February 2025

1 Introduction

How likely is it that a person is infected with COVID if they have tested positive on a COVID rapid test? Questions like this have repeatedly preoccupied us during the course of the pandemic and involve what is known as a "conditional probability." Conditional probabilities frequently pose challenges for people in our society, a striking example of which is Donald Trump's statement in October 2020 that 85% of people who wore masks nevertheless got the coronavirus (Dale, 2020; Enaganti et al., 2022). In this statement, he confused two "inverse" proportions. In fact, the Centers for Disease Control and Prevention had found in a study of 154 participants that, of those who were ill, a total of 85% said that they wore a mask either "always" or "often" over the 2 weeks prior to the beginning of their illness. Unfortunately, errors like this frequently occur (Shaughnessy, 1992).

Especially difficult in the context of conditional probabilities are Bayesian tasks (i.e., probabilistic tasks that can be solved with Bayes' formula; for an example, see the well-known *mammography problem* in Table 1), which regularly lead to wrong judgments even among experts in various fields like law or medicine (Hoffrage et al., 2000; Operskalski & Barbey, 2016; Spiegelhalter et al., 2011). In Bayesian reasoning situations, two strategies can help one to overcome typical difficulties: (1) replacing percentages by natural frequencies (e.g., saying "80 out of 100 infected people receive a positive test result" instead of "80% of the infected people receive a positive test result," see the two columns of Table 1) and (2) visualizing the statistical information, see Fig. 1 (Binder et al., 2015; Böcherer-Linder & Eichler, 2019; Gigerenzer & Hoffrage, 1995; Khan et al., 2015; McDowell et al., 2018). Despite its obvious relevance to mathematics education, research on Bayesian

Table 1 The well-known mammography problem is a typical Bayesian situation requiring the calculation of an "inverse" conditional probability

	Probability format	Frequency format
Problem	The probability of breast cancer (D) is 2% for a woman at age forty who participates in routine screening If a woman has breast cancer, the probability is 80% that she will have a positive mammogram result (T +) If a woman does not have breast cancer, the probability is 10% that she will also have a positive mammogram result What is the probability that a woman in this age group who has a positive mammogram result actually has breast cancer?	200 out of every 10 000 women at age forty who participate in a routine screening have breast cancer (D) 160 out of every 200 women with breast cancer will have a positive mammogram result (T +) 980 out of every 9 800 women without breast cancer will also have a positive mammogram result How many women with a positive mammogram result actually have breast cancer?
Correct solution	Possible solution algorithm: $P(D\|T+) = \frac{P(D\cap T+)}{P(T+)} = \frac{P(T+\|D)\cdot P(D)}{P(T+\|D)\cdot P(D)+P\left(T+\|\overline{D}\right)P(\overline{D})} = \frac{0.8\cdot 0.02}{0.8\cdot 0.02+0.1\cdot 0.98} \approx 14\%$	160 women who have breast cancer also have a positive mammogram result. 980 women do not have breast cancer but have a positive mammogram result. Therefore, (160 + 980) = 1 140 women have a positive mammogram result; of those, 160 women actually have breast cancer 160 out of 1 140

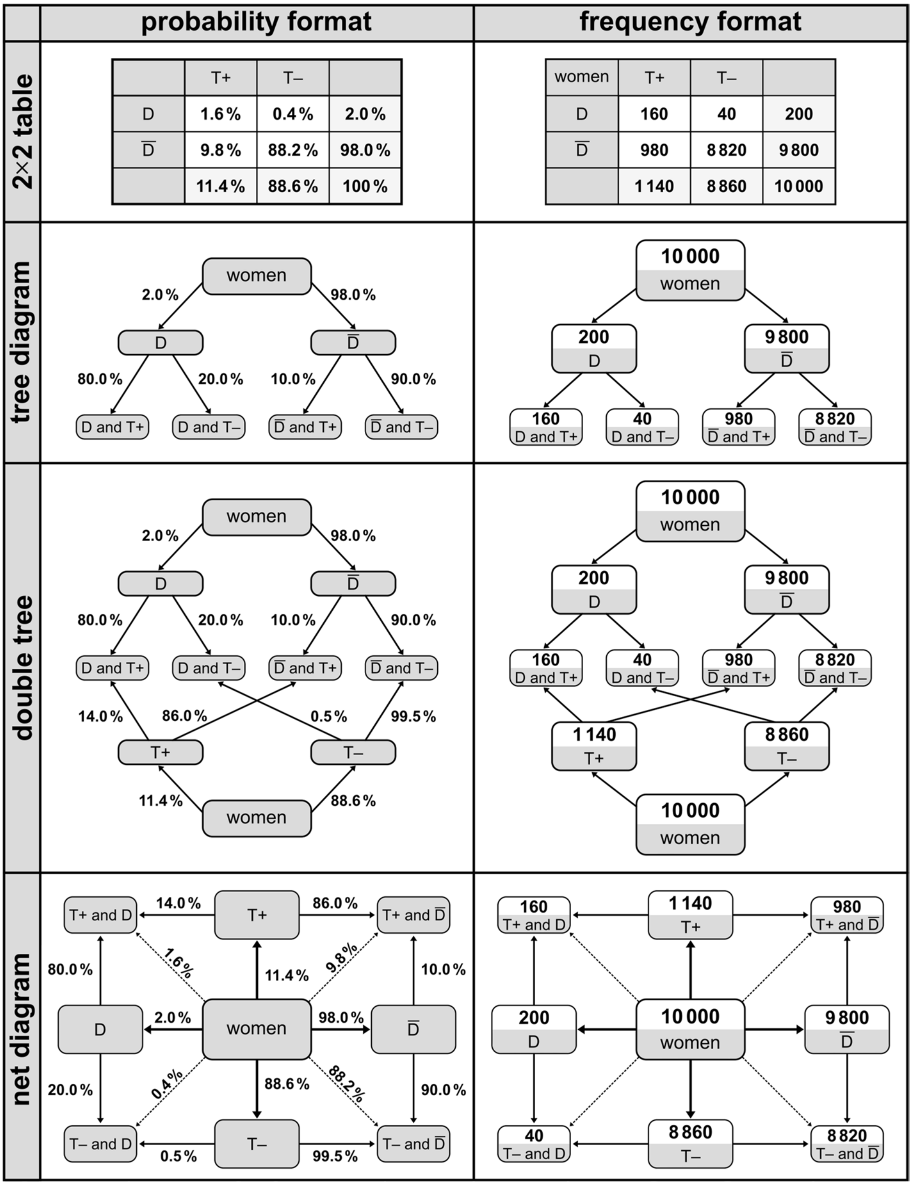

Fig. 1 Visualizations of situations with two binary events (based on the values of Table 1, D: has breast cancer, $\overline{D}$: does not have breast cancer, T+: positve mammogram result, T–: negative mammogram result). Left: probability visualizations (i.e., the numerical values are presented as percentages, decimal numbers, or fractions); right: frequency visualizations. All sixteen elementary probabilities are depicted simultaneously only in the probability net. Especially for node-branch-structures, it is possible to illustrate probabilities and frequencies simultaneously in one visualization (e.g., by superimposing the left and the right diagram)

reasoning has been predominantly conducted by cognitive psychologists (e.g., see Barbey & Sloman, 2007; McDowell & Jacobs, 2017), focusing primarily on participants' (intuitive) performance in tasks like those in Table 1.

In this paper, we want to broaden Bayesian reasoning as it is investigated so far by psychologists by implementing the perspective of mathematics education. This will be accomplished by three major shifts of focus.

First, the focus of this paper is not limited to the special case of Bayesian reasoning but covers probabilistic situations involving two binary events (i.e., health status and medical test result) in general. From a mathematical point of view, these situations feature sixteen elementary probabilities (four marginal, four joint, and eight conditional probabilities). However, most Bayesian reasoning tasks feature only four specific probabilities (three given, one asked), yet there are a lot of other probabilities that can be involved in such situations. Nevertheless, since previous research has predominantly focused only on Bayesian reasoning, this special case will play a major role in the theoretical part of this paper.

Second, in mathematics education, tasks that are related to real-world situations in general or applications of probability in particular can be interpreted in a modeling context (Gage, 2012; Kaiser & Sriraman, 2006) involving several steps that need to be successfully completed during the solving process. Although in the last 30 years a flurry of research on Bayesian reasoning has investigated the effect of information formats (e.g., probabilities vs. natural frequencies) and the power of visualizations (e.g., tree diagrams or 2×2 tables, see Fig. 1), most of this research has been based on stimuli, where the beneficial elements were already implemented by the experimenter (for a few training studies, see Büchter et al., 2022; Feufel et al., 2023; Sedlmeier & Gigerenzer, 2001; Steib et al., 2024) and therefore neglect many relevant modeling steps. For everyday teaching and learning of probability in school, however, it is indispensable that students can in the end construct visualizations on their own—even out of a real situation. Thus, it is important to *look at the entire modeling cycle* (see, e.g., Blum & Leiss, 2007, Fig. 2) and what difficulties learners face in each of the individual steps.

Third, awareness of potential errors in any step of the solving process is a crucial aspect of a teacher's professional knowledge. Thus, the research on common erroneous strategies also should be broadened to other steps of the modeling process. In dealing with "conditional information" (in the following we understand this term to mean a *conditional probability* or the corresponding *natural frequency*), first indications of typical errors and how these depend on the information format and the type of visualization being used have already been documented in studies on Bayesian reasoning. However, little is known about typical errors concerning other steps of the modeling cycle or even other inferences (e.g., requiring joint probabilities).

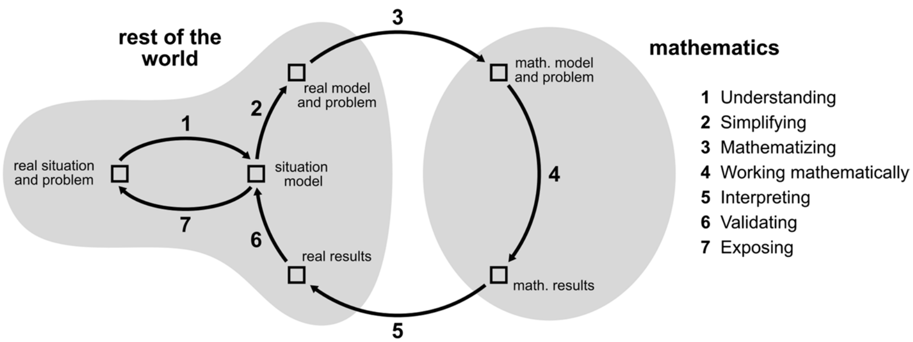

Fig. 2 The modeling cycle by Blum and Leiss (2007).

Therefore, this article pursues two goals: first, in the theoretical background, the process of solving probabilistic tasks with two binary events using a visualization is embedded in a modeling cycle. Second, one aspect is examined in more detail in an empirical study: while previous studies have mainly focused on problem solving or/and *reading* from filled visualizations, the present study looks at another step of the modeling cycle, focusing on *entering statistical information into empty visualizations* as part of peoples' ability to use (known) visualizations for mathematical modeling (i.e., in the context of probabilistic tasks but not limited to Bayesian tasks). Therefore, we examine the influence of information format and visualization type on participants' ability to fill numerical conditional or joint information (i.e., probabilities or frequencies) at the correct localizations in empty visualizations.

2 Theoretical background

In the following, probabilistic situations and corresponding tasks with two binary events are interpreted in the modeling framework (2.1) and the relevant conceptual and procedural knowledge is outlined (2.2). Next, an overview of the role and functions of visualizations in probability concerning situations with two binary events is given (2.3), and expanded by a summary of their most important characteristics (2.4). Finally, the last part gives an overview of the current state of research concerning the use of visualizations for solving probabilistic tasks with two binary events (2.5).

2.1 The modeling cycle for probabilistic tasks with two binary events

In general, modeling is seen as an important activity in mathematics and has been organized into various mathematical (e.g., Blum and Leiss, 2007) and statistical modeling cycles, such as the PPDAC cycle (Wild & Pfannkuch, 1999). Because probability models (e.g., modeling hypothetical structural relationships) are fundamental to the teaching of statistics and stochastics we specify the well-known modeling cycle by Blum and Leiss (2007), see Fig. 2, for the use of visualizations to solve probabilistic tasks with two binary events, see Fig. 3 (see also, e.g., Eichler & Vogel, 2015). In order to better illustrate the various stages of

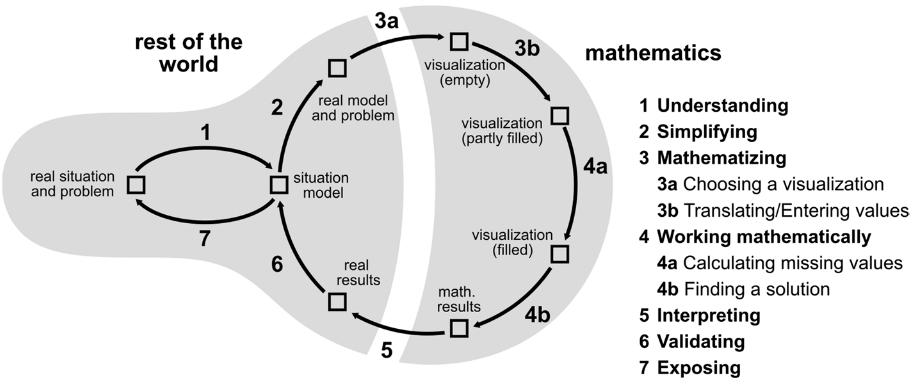

Fig. 3 The modeling cycle by Blum and Leiss (adapted, 2007), specified for probability problems with two binary events that are solved using a visualization. A concrete modeling task here could be that a marginal probability, a joint probability, and a conditional probability are given in a real modeling context (in a non-underdetermined task), and another marginal probability is to be calculated. Typical Bayesian tasks would also be conceivable here.

Table 2 The modeling cycle for solving probabilistic tasks involving two binary events based on a visualization

Step 1 **Understanding**	In reality, probabilistic situations with two binary events do not occur as neatly structured as the example for a Bayesian reasoning task in Table 1 and usually include unnecessary information, for example, information on prevalence in other age groups. Therefore, the first task for the problem solver is to construct a situation model, which means to gain a structured overview of the situation as a whole
Step 2 **Simplifying**	In order to transfer the situation model into a real model, the problem is structured further by identifying and focusing on its relevant aspects (e.g., a situation with the two binary events health status and test result; conjecture that there might be a dependency between the two events) and omitting any excess information, such as prevalence in other age groups
Step 3a **Mathematizing: Choosing a visualization**	Tasks with two binary events are often solved with the help of visualizations and this is also strongly encouraged in the classroom. In the following—also in line with the focus of this article—we look at a typical solution algorithm that uses a visualization as a heuristic tool. In this case, a suitable visualization is chosen and drawn to represent the situation. In certain cases (e.g., if the situation involves an event with more than two characteristics), the visualization needs to be adapted in order to represent the situation adequately. Success in solving (probabilistic) tasks depends largely on the type and quality of the used visualization, as well as on the students' strategic knowledge for the construction of adequate visualizations (Hembree, 1992; Rellensmann et al., 2017; Zahner & Corter, 2010), which can be seen as conceptual knowledge according to the distinction of conceptual and procedural knowledge of Hiebert and Lefevre (1986)
Step 3b **Mathematizing: Entering values**	Afterward, the given information needs to be correctly inserted into the chosen visualization, which requires procedural knowledge (e.g., how to calculate conditional probabilities from joint probabilities) and conceptual knowledge (e.g., a sufficient comprehension of subset relations, part-whole relationships, and part-part relationships, different types of probabilities and their differentiation; Dröse et al., 2022; Post & Prediger, 2022; Shaughnessy, 1992). Furthermore, knowledge on the visualization's structure and limitations is helpful, since not all probabilities can be entered into every visualization (e.g., 2×2 tables are rather unsuitable for displaying conditional probabilities; tree diagrams do not display joint probabilities, see also Sect. 2.4), which can be seen as conceptual knowledge according to the distinction of conceptual and procedural knowledge of Hiebert and Lefevre (1986)
Step 4a **Working mathematically: Determining missing values**	If needed, missing values are calculated and added to the visualization. This requires knowledge on the visualization as a symbolic representation (procedural knowledge type 1, Hiebert & Lefevre, 1986). For example, in a tree diagram, percentages on branches coming from the same node always add up to 100%. However, appropriate conceptual knowledge can also be helpful here to avoid typical errors such as confusing joint and conditional probabilities (Dröse et al., 2022) and appropriate procedural knowledge type 2, i.e., knowledge of rules and algorithms (Hiebert & Lefevre, 1986), is helpful, e.g., to calculate a conditional probability with the help of a marginal and a joint probability

Table 2 (continued)

Step 4b **Working Mathematically: Finding a solution**	After calculating the missing values, one can either simply read the desired probability from the visualization, or it needs to be calculated (e.g., a conditional probability is calculated as a fraction of two numbers in the 2×2 table), which again requires procedural knowledge in using the formula (procedural knowledge type 2) or visualizations (procedural knowledge type 1) adequately, and conceptual knowledge that prevents one from confusing different probabilities (e.g., the probability of B given A with the probability of A given B)
Step 5 **Interpreting**	After calculating a mathematical result, it has to be transferred back into the real-world scenario and is interpreted as an answer to the original question or the real model. This step requires a change of language and making sense of the numerical values: Mathematical solutions use abstract formulations, while the real-world result uses a wording similar to the situation model, for example, "the probability of being infected, given a positive test result, is 14%"
Step 6 **Validating**	The result is then used to validate the utility of the constructed model. For typical Bayesian tasks, the calculated results are often substantially lower than most people would expect them to be. For example, the positive predictive value (i.e., the probability that a woman actually has cancer if she has a positive test result) for the mammography problem is frequently overestimated (Casscells et al., 1978; Eddy, 1982). Therefore, it is necessary to check the modeling assumptions once again. This includes reviewing the visualization that was used regarding its internal coherence and concordance with the situation model
Step 7 **Exposing**	If the model is considered to be sufficient for solving the task, the results can be applied to the real-world situation, for example, by reaching a decision. Otherwise, steps 2 to 9 are run through once again with further adjustments to the model

constructing a probabilistic visualization, steps 3 ("mathematizing") and 4 ("working mathematically") are split into substeps. A short description of the individual steps can be found in Table 2. Please note that this modeling cycle depicts only one out of many possible ways to solve a probabilistic task with two binary events using a visualization. Furthermore, the steps do not necessarily have to be accomplished in this particular order and the cycle can be entered and exited at any point—depending on the complexity of the task at hand.

2.2 Conceptual and procedural knowledge for solving probabilistic tasks with two binary events

Solving probabilistic problems with two binary events requires conceptual and procedural knowledge at various points in the modeling cycle, as already outlined in Table 2 (see also Dröse et al., 2022). In the following, we refer to the distinction between conceptual knowledge and two types of procedural knowledge according to Hiebert and Lefevre (1986). For understanding probabilistic problems containing two binary events, conceptual knowledge is required to recognize and distinguish marginal, conditional, and joint information (Shaughnessy, 1992), which builds on prior mathematical knowledge of the different types of probabilities like part-whole relationships and part-part relationships (Prediger & Schink, 2009). Moreover, in order to understand what, for example, "20% out of 40%" means, it is necessary to have developed the relevant basic ideas (also called "Grundvorstellungen") of percentages or fractions. Depending on the situation and the exact wording, the phrase can mean "20%· 40% = 8% of the whole" (if it refers to a joint probability: e.g., “20% of the 40% (women) of all persons are ill”) or “20%/40% = 50% of the part” (if it refers to a conditional probability: e.g., “20% of the 40% women,” Wiesner et al., in press). This is closely linked to grasping *part-part relationships* or *part-whole relationships* in a probabilistic situation (Post & Prediger, 2022), that is, correctly identifying the two nested subsets of the sample space that make up a probability and their relationship to each other (e.g., as numerator and denominator of a fraction or as the two numbers in a natural frequency).

Relevant procedural knowledge type 1 (according to Hiebert & Lefevre, 1986, see also Sáenz, 2009) includes knowledge of the symbolic representation (and formal language) and therefore the ability to construct and—if necessary—modify appropriate visualizations for the given situation. Moreover, procedural knowledge type 2, that means, knowledge of rules and algorithms, is necessary in order to determine missing values, for example, the rules for calculating with probabilities (e.g., $P(\overline{A}) = 1 - P(A)$ or $P(A \cap B) = P(A) \cdot P(B|A)$). Additionally, language skills are crucial to decode the verbal information and interpret it mathematically and vice versa.

Visualizations play a special role for the relevant knowledge on probabilistic situations, and the conceptual knowledge and also procedural knowledge type 1 are closely related to visualizations: on the one hand, commonly used visualizations (e.g., 2×2 tables or branching structures like tree diagrams) and their characteristics can be considered to be *part of* the procedural knowledge type 1: learners need to know what information is displayed in which way in different visualizations (see 2.4). On the other hand, visualizations can be used *to build* the above-described conceptual knowledge, for example, by displaying different relationships of the several subsets of a sample space in order to foster the differentiation of the marginal, joint, and conditional probabilities. Furthermore, when creating and working with visualizations, it becomes clear whether students have already developed sufficient conceptual knowledge and procedural knowledge about the different types of probabilities. Therefore, a visualization is also a carrier of this knowledge. This results in different functions of visualizations, which will be further discussed in the following section.

2.3 The role and functions of probabilistic visualizations

Similar to the dual cognitive-communicative function of language in mathematics (Maier & Schweiger, 1999), probabilistic visualizations serve multiple purposes. First, they are considered to be a helpful *heuristic tool* that can be useful in mathematical problem solving or modeling, even for experts (Hembree, 1992; Rellensmann, 2019; Uesaka et al., 2007). In probability, for example, tree diagrams can be used to organize and enumerate a sample space (Nunes et al., 2014) and students have been reported to spontaneously create visualizations when solving probabilistic or combinatoric tasks (Zahner & Corter, 2010).

Second, probabilistic visualizations can also be used as a *medium for communication*. Although many visualizations serve *only* a communicative function (e.g., bar charts or pie charts), some predominantly heuristic visualizations can take on both functions (e.g., tree diagrams, see Spiegelhalter et al., 2011).

Third, probabilistic visualizations are also used as *teaching material* to build up conceptual knowledge about probabilistic concepts. They can help learners see connections between part-part relationships and conditional probabilities, as well as the connection between part-whole relationships and joint or marginal probabilities (Dröse et al., 2022). Furthermore, visualizations are also suitable for developing relevant conceptual knowledge so that they can enable the differentiation of conditional and joint probabilities and their conscious contrasting (Díaz & Batanero, 2009; Shaughnessy, 1992).

Taking these functions into account, the educational standards adopted by many countries see fundamental visualizations like 2×2 tables or tree diagrams as *learning objects* themselves that are therefore mandatory in coursework (ACARA, 2022; Kultusministerkonferenz, 2022; NCTM, 2023). This suggests that familiarity with common probabilistic visualizations itself can be considered to be part of the conceptual knowledge and not just a by-product of probability education. However, it is important that these visualizations are not exclusively treated as fixed structures but can also be adapted and used flexibly if necessary while still remaining their key characteristics.

In the context of probabilistic tasks with two binary events, we want to study visualizations (see Fig. 1) primarily in their function as heuristic tools that can facilitate certain steps of the solving process of these tasks (see Fig. 3). In contrast, most previous studies on (intuitive) Bayesian reasoning studied visualizations primarily in their (passive) communicative function since visualizations were usually provided completely filled and often in place of a text that presented the relevant statistical information for the following task. However, the present study does not aim to fully cover the function of heuristic tools and all steps of the modeling cycle at once. Nevertheless, it focusses on participants' ability to use their conceptual knowledge of probabilities on known and adapted visualizations.[1]

The relevant conceptual differences between the visualizations that will be used in the empirical part of this paper are outlined in the following section.

2.4 Visualizations of two binary events

In probability education, students can use various visualizations as heuristic tools to solve (modeling) problems involving two binary events. The present study focuses on 2×2 tables, double trees, and net diagrams (see Fig. 1). While 2×2 tables are quite common, double trees and net diagrams are largely unused in schools.

Probabilistic visualizations differ as to which statistical information is displayed. For two binary events, there is a total of sixteen elementary probabilities: four *marginal probabilities* $P(D)$, $P(\overline{D})$, $P(T+)$, $P(T-)$, four *joint probabilities* $P(D \cap T+)$, $P(\overline{D} \cap T+)$, $P(D \cap T-)$, $P(\overline{D} \cap T-)$, and eight *conditional probabilities* $P(D|T+)$, $P(D|T-)$, $P(\overline{D}|T+)$, $P(\overline{D}|T-)$, $P(T+|D)$, $P(T+|\overline{D})$, $P(T-|D)$, and $P(T-|\overline{D})$.[2] Most visualizations do not explicitly include all of those 16 probabilities. In many cases, a visualization focuses on *either* joint or conditional probabilities and thus neglects the other one. Moreover, some visualizations are not symmetric in the sense that a conditional probability and its reverse (e.g., $P(D|T+)$ and $P(T+|D)$) are not equally evident (e.g., tree diagrams display only one of the two possible hierarchical structures; see Fig. 1).

Table 3 presents the most important features for some visualizations in both formats. Remarkably, net diagrams (Binder et al., 2020) display all sixteen elementary probabilities, which can be seen as a structural advantage. However, students' ability to switch between visualizations, to draw connections between different visualizations, and to choose appropriate visualizations for given tasks (e.g., depending on the information that is given or that has to be calculated) are seen as key competencies summarized in the concept of representational flexibility that has been shown to be an important part in successful (flexible) mathematical problem solving in general (Acevedo Nistal et al., 2009; Novick & Hmelo, 1994; Zahner & Corter, 2010).

[1] Double trees and net diagrams are usually not (yet) treated in schools but follow the same rules for entering probabilities or frequencies as tree diagrams as another well-known node-branch-structure.

[2] Besides these sixteen elementary probabilities, there are $P(\varnothing)$, $P(\Omega)$, and the probabilities of unions of subsets (e.g., $P(D \cup T+)$).

Table 3 Features of the 2×2 table, tree diagram, double tree, and net diagram in both formats, which can be used as heuristic tools for modeling situations with two binary events

Visualization	Format	Marginal probabilities/ frequencies can be displayed directly	Joint probabilities/ frequencies can be displayed directly	Conditional probabilities/frequencies can be displayed directly
2×2 table	**Probabilities**	✓	✓	no
	Frequencies	✓	✓	✓
Tree diagram	**Probabilities**	Only 2 out of 4	no	Only 4 out of 8
	Frequencies	Only 2 out of 4	✓	Only 4 out of 8
Double tree	**Probabilities**	✓	no	✓
	Frequencies	✓	✓	✓
Net diagram	**Probabilities**	✓	✓	✓
	Frequencies	✓	✓	✓

2.5 Previous research on the use of visualizations for probabilistic tasks with two binary events

Previous research has primarily focused on the special case of Bayesian reasoning but has nevertheless generated important insights into the effects of information formats and visualizations on participants' performance and typical errors in solving probabilistic tasks with two binary events in general. First, Bayesian tasks are solved correctly more often if the task is presented in natural frequencies instead of probabilities (e.g., "80 out of 100 infected people receive a positive test result" instead of "80% of the infected people receive a positive test result", *frequency effect*, Gigerenzer & Hoffrage, 1995; McDowell & Jacobs, 2017). Second, presenting the task using a completely filled visualization instead of or next to a textual representation can also facilitate solving Bayesian tasks (Binder et al., 2015; Böcherer-Linder & Eichler, 2019; Gigerenzer & Hoffrage, 1995; Khan et al., 2015; McDowell et al., 2018). Furthermore, visualizations seem to be of greater benefit if displayed in the frequency format rather than in the probability format (see Fig. 1, Binder et al., 2015). Yet, recent studies have shown that the frequency effect (for visualizations) does not hold if the task calls for joint instead of conditional information (Binder et al., 2020; Stegmüller et al., 2024).

The following section describes typical errors that are known from previous studies. These are mainly observable error patterns that may be based on misconceptions. It has to be noted that an observed error pattern or a correct answer does not indicate whether or not the student has a general misconception (Riccomini, 2005).

2.5.1 Typical errors in solving tasks including conditional probabilities

Various errors for dealing with conditional information, especially in the context of Bayesian reasoning, will be summarized and explained using the mammography problem (see Table 1) in the following. In most of the studies cited below, the visualizations were already completely filled and participants' ability in extracting statistical information adequately was examined.

One of the most frequently observed errors (an overview is given in Table 4) is the so-called *joint occurrence error*, where a conditional probability is wrongly interpreted as a joint probability (e.g., $P(D|T+)$ is interpreted as $P(D \cap T+)$). This error pattern may be based on the following misconception: In terms of part-whole relationships, the part is hereby recognized correctly (i.e., the numerator of the corresponding fraction or the first number in the natural frequency is identified correctly as $D \cap T+$), whereas the whole (i.e., the superset or the second number) is erroneously interpreted as the whole sample space Ω instead of the subset T+. A lack of conceptual knowledge in regard to part-whole relationships therefore contributes to this error pattern. In the example given in Table 1, the joint occurrence error would lead to the answer 1.6% or 160 out of 10 000.

Another common error is the *Fisherian*, subsequently called *inverse error*: A conditional probability is mistaken for its inverse conditional probability, that means, $P(D|T+)$ is confused with $P(T+|D)$ (in the mammography problem, this strategy yields 80% instead of 14%). In this case, the part is again identified correctly, but the corresponding whole is chosen wrongly (i.e., D instead of T+; Falk, 1986; Gigerenzer & Hoffrage, 1995; Kahneman et al., 1982). This error pattern can be caused by a lack of conceptual knowledge regarding the part-whole relationship but can also be facilitated by a lack of knowledge on how to verbally formulate different types of probabilities adequately. Erroneously assuming that the two events D and T+ are independent (i.e., $P(D|T+) = P(D)$) is called the *base-rate-only* error. Similarly, the *evidence-only* error focuses only on the conditional event, which means that $P(D|T+)$ is confused with $P(T+)$. Both of these errors result in a wrong interpretation of part and whole, leading to the answers 2% or 11.4%, respectively. Finally, ignoring the intersecting structure of the two events and therefore misinterpreting the part-whole relationship leads to the *pre-Bayes* error, where the probability $P(D|T+)$ is mistakenly expressed as $|D|/|T+|$ (or $|D|$ out of $|T+|$), yielding $200/1140 \approx 17.5\%$. This strategy uses a correct whole but a false part.

Table 4 Correct solution and typical error patterns for Bayesian tasks (Binder et al., 2020; Bruckmaier et al., 2019; Díaz & Batanero, 2009; Eichler & Böcherer-Linder, 2018; Gigerenzer & Hoffrage, 1995; Steckelberg et al., 2004; Woike et al., 2023; Zhu & Gigerenzer, 2006)

	Probabilities[a]	Frequencies[a]
Correct solution (Bayesian)	$\frac{P(D)\cdot P(T+\|D)}{P(D)\cdot P(T+\|D)+(1-P(D))\cdot P(T+\|\overline{D})}$	$\|T+\cap D\|$ out of $\|T+\|$
Error patterns		
Joint occurrence	$P(D) \cdot P(T+\|D)$	$\|T+\cap D\|$ out of $\|\Omega\|$
Inverse error	$P(T+\|D)$	$\|T+\cap D\|$ out of $\|D\|$
Base-rate-only	$P(D)$	$\|D\|$ out of $\|\Omega\|$
Evidence-only	$P(D)\cdot P(T+\|D) + (1-P(D))\cdot P(T+\|\overline{D})$	$\|T+\|$ out of $\|\Omega\|$
Likelihood-subtraction	$P(T+\|D) - P(T+\|\overline{D})$	Not applicable
False-alarm-complement	$1 - P(T+\|\overline{D})$	$\|T-\cap\overline{D}\|$ out of $\|\overline{D}\|$
50%-rule	50%	Not applicable
Missing base-rate[b]	$\frac{P(T+\|D)}{P(T+\|D)+P\left(T+\|\overline{D}\right)}$	Not applicable

[a]Probabilities and cardinalities are labeled according to Fig. 4. We consider Ω to be the set containing all individual people of the (imaginary) sample, for example, $|\Omega| = 10\ 000$ (people). Therefore, the cardinality of the subsets (e.g., $|D|$) yields the expected frequency with which the corresponding event is going to occur; [b] "missing base-rate" is similar to the Bayesian formula but missing the so-called base-rate P(D)

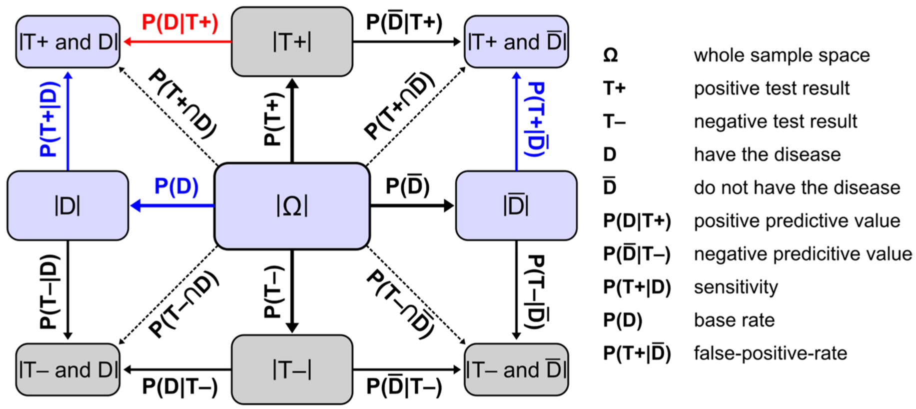

Fig. 4 Schematic net diagram for the two events *test result* (positive/negative) and *health status* (has the disease/does not have the disease), representing the sixteen elementary probabilities, as used in Table 4. Probabilities/frequencies that are given in a typical Bayesian inference task are blue; the probability that has to be calculated is red

Further errors are shown in Table 4. Since only the net diagram can display all relevant probabilities, it is used (Fig. 4) to explain the letters in Table 4.

2.5.2 Typical errors in solving tasks that include joint inferences

While typical errors involving conditional probabilities are well-known via research on Bayesian reasoning, errors in calculating joint probabilities have been studied less frequently (for an exception, see, e.g., Binder et al., 2020). To illustrate these errors, we consider the visualizations in Fig. 1 together with the question, “What is P(T+∩D)?” as an example of a joint inference task. In that case, the correct solution is 1.6%.

In analogy with the joint-occurrence error in Bayesian reasoning, a common mistake in handling joint probabilities is to not identify the whole as the entire sampling space Ω but as a smaller subset, in that way confusing joint and conditional probabilities, leading to an error which we call *conditional error*, where P(D∩T+) is confused with P(T+|D) or P(D|T+), that is, 14% or 80%, respectively, in the example in Fig. 1 (see also Table 5).

Another mistake occasionally seen—the *independence error*—originates from the erroneous use of a well-known formula: If and only if the events D and T+ are independent, P(D∩T+) = P(D)·P(T+) holds true. If insufficient conceptual knowledge of stochastic independence has been built up, the formula is sometimes misused when the joint probabilities are calculated from two not stochastically independent events. In the example, this strategy yields 2.0%·11.4% ≈ 0.2%.

In general, for both joint and conditional probabilities, one or more negations might sometimes be misread from the text, leading to the *negation error* where, for example, P($\overline{D}$|T+) is confused with P(D|T+), or P($\overline{D}$∩T+) is confused with P(D∩T+). Further errors are shown in Table 5.

Table 5 Correct solution and typical error patterns for joint inferences (Binder et al., 2020)

	Probabilities[a]	Frequencies[a]
Correct solution	P(D)·P(T+|D) or P(T+)·P(D|T+)	|T + and D| out of |Ω|
Error patterns		
Conditional error	P(T+|D) or P(D|T+)	|T + and D| out of |D| or |T + and D| out of |T +|
Independence error	P(D)·P(T+)	Not applicable
Negation error	e.g., P(T − ∩D)	e.g., |T− and D| out of |Ω|
Double joint probability	P(D)·P(T+|D)+ P(T+) · P(D|T+)	Not applicable

[a]For explanation of the letters, see Fig. 4

2.5.3 The influence of information format and visualization on typical errors

As briefly discussed earlier, visualizations differ in the way they present probabilistic information. As a result, some visualizations are particularly susceptible to typical or unique error patterns.

A 2×2 table can provoke confusion of probabilities since its inner cells serve multiple purposes: On the one hand, each of these cells directly displays one joint probability, while on the other hand, each cell can be used to determine two conditional probabilities in combination with a corresponding marginal cell in the same row or column (Batanero et al., 1996; Roca & Batanero, 2006). This seems especially challenging for the calculation of conditional information in the probability format with percentages, possibly because dividing two percentages by each other in order to get yet another percentage is a rather unfamiliar operation and requires a thorough understanding of the part-part model (note that the calculation of conditional information from 2×2 tables is independent of the information format, i.e., in both formats, the numbers from the same cells have to be divided).

Despite their straightforward structure, tree diagrams have been reported to be frequently misconstructed and/or misinterpreted, for example, in the solving of combination problems (Lamanna et al., 2022) or sequential probabilistic problems (Awuah & Ogbonnaya, 2020). However, misconstruction can also occur in situations with two binary events, especially if conditional probabilities need to be calculated, since tree diagrams focus only on one "conditional direction."

Considering the efforts that are being made in statistics education to foster students' modeling competencies (Biehler et al., 2017; Pfannkuch et al., 2018; Zapata-Cardona, 2018) and their high universal relevance (e.g., in the recent Covid pandemic: "What is the probability that I am infected if I have a positive test result?"), it is remarkable that the vast majority of empirical research for the case of visualizations for probabilistic situations with two binary events has focused almost exclusively on step 4b of the modeling cycle (reading completely filled visualizations), in particular concerning Bayesian tasks (see Fig. 5). However, when confronted with probabilistic situations in real life or in school book tasks, visualizations are not provided. Therefore, constructing a visualization and entering numerical values is a crucial aspect of

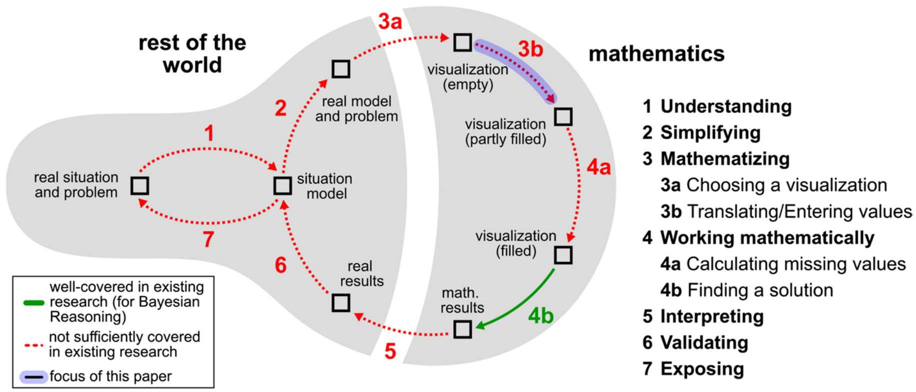

Fig. 5 The modeling cycle by Blum and Leiss (adapted, 2007), color-coded to display current state of research concerning, e.g., Bayesian reasoning

solving (but also for understanding/structuring) a probabilistic task. Although there have been isolated efforts to cover this area, the few studies focusing on earlier steps in the modeling cycle have been rather small and did not further investigate erroneous strategies (Bobek & Corter, 2010; Friederichs et al., 2014; Zahner & Corter, 2010). Therefore, the question remains open as to whether participants, who, e.g., decide to solve the task of Table 1 with the help of a visualization, are able to put numbers such as that given in the problem (Table 1) at the right places in visualizations (e.g., the ones displayed in Fig. 1, if they are still empty).

The intention of the present study is to take one step back from the usual focus (i.e., green arrow in Fig. 5) and examine students' abilities and difficulties when using visualizations as heuristic tools. The focus of the empirical study is on typical errors made in entering a given probability or natural frequency into an empty visualization.

3 Research questions and hypotheses

The current study investigates one specific aspect of students' ability in using visualizations as heuristic tools. The focus is on students' performance and typical error patterns in entering conditional or joint probabilities, or corresponding frequencies, into empty visualizations (2×2 tables, double trees, and net diagrams). The research questions (RQ), with hypotheses (H), are summarized in Table 6.

Table 6 Research questions with hypotheses

Research question	Hypotheses
RQ1: Does the format of information (probabilities vs. frequencies) and the given empty visualization affect participants' performance in entering conditional and joint information (i.e., conditional and joint probabilities, or the corresponding natural frequencies)?	**H1a:** The percentage of correct entries will be higher if the information is given in frequencies rather than probabilities (both for conditional and joint information) because we expect the format effect to hold for entering information into visualizations as well (for joint probabilities; however, the findings of Binder et al. 2020 suggest that we may not find a systematic format effect here) **H1b:** Net diagrams and double trees will lead to better performance in entering conditional *probabilities* compared to 2×2 tables, since conditional probabilities cannot be directly entered in 2×2 tables **H1c:** Net diagrams and 2×2 tables will lead to higher performance in entering joint *probabilities* compared to double trees because double trees do not directly display joint probabilities
RQ2: Does the format of information (probabilities vs. frequencies) and the given empty visualization affect typical participants' error patterns in entering conditional and joint information (i.e., conditional and joint probabilities or the corresponding natural frequencies)?	**H2a:** In general, we expect to observe analog typical error patterns in filling out visualizations that are already known from research on solving Bayesian reasoning tasks with the help of purely textual versions or with the help of completely filled visualizations. However, previous empirical findings suggest that errors in reading from visualizations can differ from errors in filling out visualizations (Cox, 1997) **H2b:** Furthermore, we expect to observe new error patterns: (1) filling in the given information in multiple places on the visualization (one of them correct and the other one incorrect), (2) just filling in the set or the subset but not both pieces of information from the natural frequency into the visualization, or (3) errors that are related to the formulated negation (e.g., test negative) **H2c:** Compared to all visualizations, the joint occurrence error will appear most often in the entering of conditional probabilities into 2×2 tables, since 2×2 tables do not directly display conditional but joint probabilities

4 Methods

4.1 Design of the study

In order to examine students' ability to use empty visualizations as heuristic tools and fill them correctly, a paper-and-pencil test was carried out using the introductory stories of two typical Bayesian tasks, the mammography problem and a short version of the economics problem (Ajzen, 1977). Participants were asked in each problem to enter one specific given numerical information into provided empty visualizations (see Fig. 6). Note that—in contrast to typical (Bayesian) reasoning problems—for each task, only one piece of statistical information was given that had to be placed into a visualization (i.e., it was not necessary to infer a, e.g., positive predictive value).

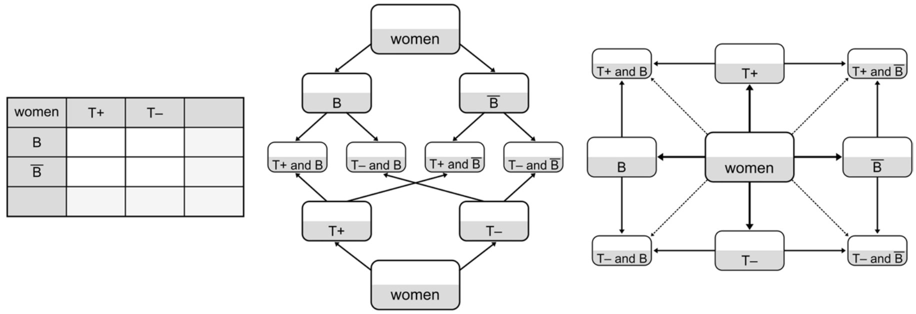

Fig. 6 Empty visualizations (2×2 table, double tree, net diagram) for the mammography context. Visualizations for the economics context were structurally equivalent but were labeled with "students" instead of "women" and E (economics course) and C (career-oriented) instead of B and T+

The design of the study includes two factors of interest and one factor that is not of interest, resulting in a 2 × 3 × 2 design (see also Table 7):

- Factor 1: Format of information—probabilities vs. frequencies
- Factor 2: Visualization—2×2 table vs. double tree vs. net diagram
- Factor 3 (not a factor of interest): Context—mammography vs. economics

In total, 24 tasks were implemented (see Table 7), and each participant completed one testlet that included four of these tasks in the following way: Tasks 1 and 2 had the same format of information, context, and visualization but asked for entering a different *type of information* (e.g., task 1 conditional information and task 2 joint information, or vice versa). Tasks 3 and 4 also differed from each other only in the type of information that had to be

Table 7 Design of the 24 used problem versions

Format of information	**Context**	**Type of information that had to be entered**	
		Conditional information	**Joint information**
Probabilities	**Mammography problem**	2×2 table	2×2 table
		Double tree	Double tree
		Net diagram	Net diagram
	Economics problem	2×2 table	2×2 table
		Double tree	Double tree
		Net diagram	Net diagram
Natural frequencies	**Mammography problem**	2×2 table	2×2 table
		Double tree	Double tree
		Net diagram	Net diagram
	Economics problem	2×2 table	2×2 table
		Double tree	Double tree
		Net diagram	Net diagram

Each participant was given a testlet containing two out of the 12 pairs of tasks, which address one conditional and one joint information (the 12 pairs are represented as 12 rows in the table). The two pairs of tasks have a different context, format, and visualization. Each of the (in sum) four individual tasks was presented on an individual sheet of paper. Two different contexts were used, in order to allow each participant to answer two different pairs of questions instead of just one entered, but had a different context, format of information, and visualization as the first two tasks. If, for instance, the mammography context appeared in the probability format with a 2×2 table in the first two tasks of a particular testlet, the remaining two tasks of the testlet featured the economics context in the natural frequency format with a double tree or a net diagram. The order of context, information format, visualization, and type of information were varied systematically between the testlets so that every possible combination was covered.

4.2 Materials

Each task began with an introductory story (see Table 8); after that, one of the three different kinds of empty visualizations, presenting only the labeling of the events (but no

Table 8 Task formulations

Context	Mammography		Economics	
Introductory story	Imagine you are a reporter for a women's magazine and you want to write an article about breast cancer. As part of your research, you focus on mammography as an indicator of breast cancer. You are especially interested in the question of what it means when a woman has a positive result (which indicates breast cancer) in such a medical test. A woman can have breast cancer (B) or not ($\overline{B}$). Furthermore, she can get a positive (T+) or a negative (T–) result. All women in this exercise participate in routine screenings		Imagine you are interested in whether career-oriented students are more likely to attend an economics course. The school psychological service evaluates the correlations between stated goals and choice of courses for you. A student can be career-oriented (C) or not ($\overline{C}$). Furthermore, a student can choose an economics course (E) or not ($\overline{E}$). All students in this exercise participated in the evaluation	
Visualization	2×2 table, double tree, or net diagram		2×2 table, double tree, or net diagram	
Information format	**Natural frequency version**	**Probability version**	**Natural frequency version**	**Probability version**
Instruction	Please enter the following two **bold** printed numbers in the correct nodes/fields in the following visualization, or check [x] that the numbers cannot be entered because there is no fitting node/field	Please enter the following **bold** printed probability in the correct branch/field in the following visualization, or check [x] that the probability cannot be entered because there is no fitting branch/field	Please enter the following two **bold** printed numbers in the correct nodes/fields in the following visualization, or check [x] that the numbers cannot be entered because there is no fitting node/field	Please enter the following **bold** printed probability in the correct branch/field in the following visualization, or check [x] that the probability cannot be entered because there is no fitting branch/field
Conditional information: conditional probability or corresponding natural frequency	Out of **8 860** women who participate in routine screening and get a negative result, **40** do have breast cancer	The probability that a woman who gets a negative result has breast cancer is **0.5%**	Out of **387** students who are not career-oriented, **115** choose an economics course	The probability that a student who is not career-oriented chooses an economics course is **29.7%**
Joint information: joint probability or corresponding natural frequency	**160** out of **10 000** women have breast cancer *and* get a positive result	The probability that a woman has breast cancer *and* gets a positive result is **1.6%**	**205** out of **1 000** students are career-oriented *and* choose an economics course	The probability that a student is career-oriented *and* chooses an economics course is **20.5%**

numerical values) was given (see Fig. 6). The structure was already provided (which is important because double trees and net diagrams were new to the students), and the statistical information had to be entered in the appropriate place. For this purpose, participants got one piece of information (a conditional probability or the corresponding two absolute frequencies, or a joint probability or the corresponding two absolute frequencies; see Table 8) that they had to enter in the given visualization. They could also check a box that it was not possible to enter the given information, which was sometimes the correct answer (see also Table 8 for task formulations).

4.3 Participants

The study was carried out in late 2020 and early 2021 and therefore was affected by lockdowns due to COVID-19. In order to obtain a sufficient sample size, $N = 172$ participants were recruited both from a university ($N = 52$) and secondary-school (six classes from a German *Gymnasium*, $N = 120$) in Bavaria. The university students were at the beginning of their studies in Biology and had not taken any courses in mathematics or statistics in university. Out of the 52 university students, 38 were female and 14 male. Their ages ranged from 17 to 60, with an average of 20.73 (SD = 6.13). Out of the 120 secondary-school students, 45 were female and 75 male. They were in Grades 11 and 12, and their ages ranged from 16 to 19, with an average of 17.09 (SD = 0.89). Since conditional as well as joint probabilities are part of the Bavarian curriculum in Grades 10 and 11, all school students had encountered this topic in their last school year, but they had not dealt with it in their current school year. Therefore, all students were familiar with joint and conditional probabilities, as well as with 2×2 tables containing probabilities and frequencies and tree diagrams containing probabilities, but not with tree diagrams containing absolute frequencies in their nodes, double trees, or net diagrams.

The testlets were distributed randomly to the participants, resulting in similar group sizes, as can be seen in Table 9.

Participants were asked to state their gender, age, and math grade in their last school report card.

The study was carried out in accordance with the University Research Ethics Standards. The students were informed that their participation was voluntary and anonymity was guaranteed. The participants provided their written informed consent to participate.

Table 9 Group sizes for the different visualizations, formats, and types of information (across both contexts)

Format	Type of information	2×2 table	Double tree	Net diagram	Sum
Probabilities	Conditional	57	56	59	172
	Joint	57	56	59	172
Frequencies	Conditional	58	56	58	172
	Joint	58	56	58	172

4.4 Coding

4.4.1 Conditional probability

In the probability versions of the task, the response was coded as correct if the probability was entered at the correct branch in the double tree and net diagram or if the participant checked the box that the given conditional probability could not be entered in the 2×2 table. In the frequency versions, the response was classified as correct if both absolute numbers were entered at the correct position in the given visualization.

4.4.2 Joint probability

The entry of the given joint probability was coded as correct if the given probability was entered at the correct branch of the net diagram or cell of the 2×2 table or if the participant checked the box that the probability could not be entered in the double tree. Again, in the frequency version, the response was classified as correct if both absolute numbers were entered correctly.

4.4.3 Specific cases

The response was coded as wrong if given probabilities or frequencies were entered multiple times (exception: the magnitude of the population was entered in both corresponding nodes in the frequency double tree, compare Fig. 1). The response was also classified as incorrect if the entry of the given conditional probability at the crossing branches of the double tree could not be clearly allocated to one of the branches. This occurred only once in the study.

4.4.4 Coding of the error patterns

Coding of the typical error patterns was based on a deductively derived coding scheme, based on the well-known error patterns in reading information from diagrams (see theoretical background), complemented by inductively derived categories that occurred during the coding process (Mayring, 2014). A code manual was developed to outline the typical errors that we expected would occur in the study. Two raters coded 20% of all (conditional and joint) entries independently according to this code manual (see Online Resources 1 and 2). Since all of the correct entries were rated in agreement with each other, and the typical errors were classified identically in 99.3% of all cases, the remaining entries were rated only by one coder.

In order to reduce the number of different error categories and present the results clearly, only error patterns that were committed at least five times in any of the six versions were accepted to be proper error pattern. Less frequent errors were summarized in a category "other uniquely classifiable errors." Nevertheless, the inverse error is additionally listed because this is a well-known error and frequently occurred in solving pure text variants of Bayesian tasks.

Sometimes, it was possible for one answer to be assigned to two or more error categories, for example, confusing joint with conditional probabilities, and also dropping a negation (e.g., P(D∩T+) is confused with P(D|T−)). Some errors could not be assigned to any category and were therefore labeled as "error unknown."

4.5 Statistical model

In order to statistically compare the effects of information format and type of visualization, we estimated generalized linear mixed models (GLMM) with a logit link function to predict performance in entering conditional and joint information. Because of the dichotomous dependent variable (0 = incorrect, 1 = correct solution), we refrained from calculating a linear regression. Furthermore, we decided for a mixed analysis and against a, for instance, logistic regression, due to our between-within-subject design since each participant solved several tasks. To take this aspect into account, we modeled the participants' ID as a random factor. We specified the natural frequency version with a net diagram as the reference category and included the possible explanatory factors "probabilities," "2×2 table," and "double tree" via dummy coding. Furthermore, we included the interaction terms (probabilities × 2×2 table) and (probabilities × double tree) and implemented the mathematics grade from the last school report card into the model. Please note that grades in Germany range from 1 to 6, with 1 being the best possible grade.

5 Results

5.1 Participants' performance in entering conditional and joint information

5.1.1 Conditional information

When students are required to use empty visualizations as a heuristic tool and fill in the given statistical information in the appropriate places, participants entered frequencies correctly more often into 2×2 tables and net diagrams than conditional probabilities (Fig. 7). For double trees, however, both formats were entered almost equally often correctly. The highest performance rate was found for entering frequencies into the net diagram, and the lowest for stating that conditional probabilities cannot be entered into 2×2 tables.

As can be seen from Table 10, the (unstandardized) regression coefficient for entering probabilities instead of natural frequencies was significant and negative as hypothesized (H1a), which means the students were able to enter two absolute frequencies in the net diagram more easily than one conditional probability. Presenting a 2×2 table or a double tree instead of a net diagram only leads to a descriptive but not a significantly worse performance of participants (H1b). Furthermore, there is a significant interaction effect between information format and 2×2 table, whereas the interaction between information format and double tree is not significant. This means that in the 2×2 table, the information format has an even stronger influence on the participants' performance than it does in the net diagram. Furthermore, a student's mathematics grade is a significant predictor for correct entry of the numbers. Please note that the participants were familiar with 2×2 tables but not with

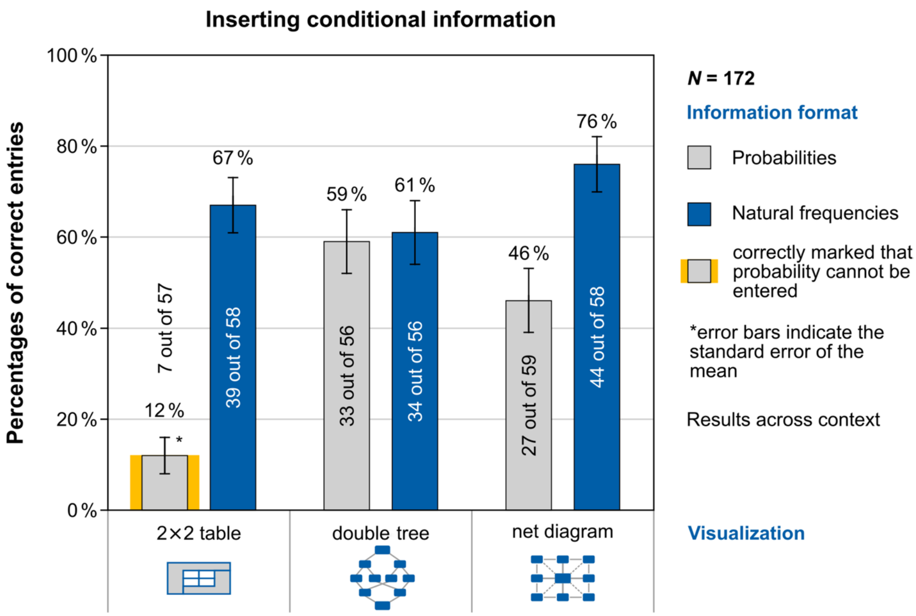

Fig. 7 Percentages of correct entries of conditional information, separated by information format and visualization type (across both contexts)

Table 10 Parameter estimates from the generalized linear mixed model for participants' performance in entering conditional information

Covariates	Estimate	SE	z	p
Intercept	2.37	0.49	4.84	**< 0.001**
Probabilities	− 1.18	0.42	− 2.79	**< 0.01**
2×2 table	− 0.26	0.43	− 0.59	0.55
Double tree	− 0.58	0.43	− 1.34	0.18
2×2 table × probabilities	− 1.84	0.68	− 2.71	**< 0.01**
Double tree × probabilities	1.05	0.59	1.80	0.07
Mathematics grade	− 0.42	0.12	− 3.55	**< 0.001**

$R^2_{\text{marginal}} = 28.7\%$, $R^2_{\text{conditional}} = 30.2\%$

double trees and net diagrams. Additionally, participants might prefer to enter a value over stating that a value cannot be entered, if they are unsure.

5.1.2 Joint information

For the double tree and net diagram, frequencies were entered correctly more often than joint probabilities (Fig. 8). However, almost all participants were able to enter joint probabilities correctly into 2×2 tables. The lowest performance was found for responding that joint probabilities cannot be entered into double trees.

As can be seen from Table 11, the (unstandardized) regression coefficient for probabilities was significantly negative, which means that probabilities were entered correctly into the net

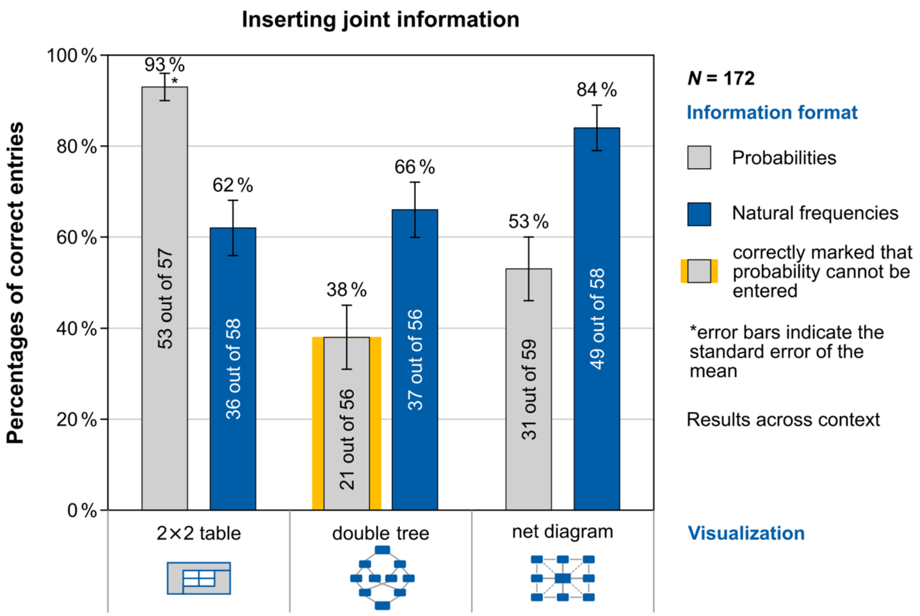

Fig. 8 Percentages of correct entries of joint information, separated by information format and visualization type (across both contexts)

diagram significantly less often than natural frequencies (H1a). Also, presenting a 2×2 table or a double tree led to a significant negative regression coefficient, which demonstrates an advantage of the frequency net (H1c). There is no significant interaction effect between information format and double trees, which means that the influence of the information format within the visualization is comparable in net diagrams and double trees. However, there is a significant positive interaction effect between information format and 2×2 table, which means that if 2×2 tables are considered instead of net diagrams, this interaction effect significantly counteracts the negative effect of the information format in net diagrams (here, probability 2×2 tables help better than frequency 2×2 tables, whereas, in net diagrams, it is the other way around, contrary to H1a). This means there is no general "frequency effect" for entering *joint* information, which has been repeatedly observed in studies focusing on *conditional* probabilities (Gigerenzer & Hoffrage, 1995; McDowell & Jacobs, 2017). Furthermore,

Table 11 Parameter estimates from the generalized linear mixed model for participants' performance in entering joint information

Covariates	Estimate	SE	z	p
Intercept	2.85	0.58	4.91	**< 0.001**
Probabilities	− 1.61	0.49	− 3.30	**< 0.001**
2×2 table	− 1.11	0.48	− 2.30	**0.02**
Double tree	− 1.03	0.49	− 2.09	**0.04**
2×2 table × probabilities	3.70	0.81	4.56	**< 0.001**
Double tree × probabilities	0.33	0.65	0.51	0.61
Mathematics grade	− 0.37	0.13	− 2.81	**< 0.01**

$R^2_{\text{marginal}} = 28.3\%$, $R^2_{\text{conditional}} = 34.6\%$

the mathematics grade in the last school report card is again a significant predictor of correct entry of numbers.

We found no significant effects regarding gender, the order of the task, subgroup (university vs. school students), and the context of the task when including them into the models. Furthermore, implementing these factors did not change the main findings for predicting students' ability in filling empty diagrams with conditional or joint information found by the GLMMs described above.

5.2 Typical error patterns when entering conditional and joint information

5.2.1 Conditional information

Table 12 separately lists typical error patterns encountered by information format and visualization. First, typical errors, which are known from research on reading information from visualizations, can be observed (confirming H2a), and some new errors occur in the filling in of information in the visualizations used as heuristic tools (confirming H2b). Furthermore, it is striking that the joint occurrence error was typical for the probability 2×2 table (84%) but dropped to almost zero in the frequency 2×2 table (confirming H2c). Such a large performance difference between probability and natural frequency versions when 2×2 tables were

Table 12 Percentages (and frequencies) of typical error patterns in entering conditional probabilities (or corresponding frequencies) in the different visualizations

	2×2 table		**double tree**		**net diagram**	
	Probabilities	Frequencies	Probabilities	Frequencies	Probabilities	Frequencies
Correct solution	**12 %** (7)	**67 %** (39)	**59 %** (33)	**61 %** (34)	**46 %** (27)	**76 %** (44)
Error patterns						
Joint occurrence	**84 %** (48)	**3 %** (2)	**0 %** (0)	**11 %** (6)	**32 %** (19)	**3 %** (2)
Inverse error	**0 %** (0)	**3 %** (2)	**5 %** (3)	**4 %** (2)	**0 %** (0)	**0 %** (0)
Incomplete (set or subset missing)	**n.a.**[a]	**5 %** (3)	**n.a.**	**9 %** (5)	**n.a.**	**3 %** (2)
Multiple answers	**0 %** (0)	**0 %** (0)	**11 %** (6)	**0 %** (0)	**3 %** (2)	**0 %** (0)
Negation error	**5 %** (3)	**9 %** (5)	**4 %** (2)	**7 %** (4)	**2 %** (1)	**2 %** (1)
Mistakenly checked "not possible"	**n.a.**	**10 %** (6)	**18 %** (10)	**9 %** (5)	**14 %** (8)	**9 %** (5)
Other uniquely classifiable errors	**0 %** (0)	**2 %** (1)	**4 %** (2)	**7 %** (4)	**0 %** (0)	**10 %** (6)
Error unknown	**4 %** (2)	**2 %** (1)	**4 %** (2)	**0 %** (0)	**3 %** (2)	**0 %** (0)

The column sums may add up to a number above 100% because in some cases two errors occurred at the same time (see Sect. 4.4). However, these combinations occurred in an unsystematic way and are therefore not reported here but can be seen in Online Resource 3. Errors marked in blue cannot occur when reading from visualizations. [a]Not applicable used may be due to the fact that many uncertain participants preferred to enter the given probability instead of checking the box that this was not possible. Also, with the probability net, where there are branches for the joint probability, 32% of the participants committed this error as opposed to 3% of those using the frequency net. In the probability double tree, where such a branch is not even presented, this error cannot be made. Instead, participants often either entered the probability in several branches on the double tree or stated that the conditional probability could not be displayed within this visualization. However, in frequency double trees, more than a quarter of the participants either committed the joint occurrence error, failed to enter a complete frequency pair (*incomplete*), or stated that the information could not be entered (*mistakenly checked "not possible"*).

5.2.2 Joint information

Since little is known thus far about error patterns regarding joint probabilities (for an exception, see Binder et al., 2020; Stegmüller et al., 2024), this analysis looking for typical errors while using the different heuristic tools was rather exploratory.

Table 13 lists the typical error patterns separately by information format and visualization. In the probability format, most participants who were not able to provide the correct

Table 13 Percentages (and frequencies) of typical error patterns in entering joint information in the different visualizations

	2×2 table		**double tree**		**net diagram**	
	Probabilities	Frequencies	Probabilities	Frequencies	Probabilities	Frequencies
Correct solution	**93 %** (53)	**62 %** (36)	**38 %** (21)	**66 %** (37)	**53 %** (31)	**84 %** (49)
Error patterns						
Conditional error	**0 %** (0)	**7 %** (4)	**50 %** (28)	**0 %** (0)	**22 %** (13)	**0 %** (0)
Marginal	**0 %** (0)	**2 %** (1)	**2 %** (1)	**18 %** (10)	**0 %** (0)	**3 %** (2)
Incomplete (set or subset missing)	**n.a.**[a]	**16 %** (9)	**n.a.**	**4 %** (2)	**n.a.**	**5 %** (3)
Multiple answers	**0 %** (0)	**2 %** (1)	**2%** (1)	**4 %** (2)	**0 %** (0)	**2 %** (1)
Negation error	**2 %** (1)	**2 %** (1)	**2 %** (1)	**2 %** (1)	**7 %** (4)	**2 %** (1)
Mistakenly checked "not possible"	**5 %** (3)	**14 %** (8)	**n.a.**	**4 %** (2)	**19 %** (11)	**2 %** (1)
Other uniquely classifiable errors	**0 %** (0)	**2 %** (1)	**5 %** (3)	**2 %** (1)	**2 %** (1)	**2 %** (1)
Error unknown	**0 %** (0)	**2 %** (1)	**2 %** (1)	**2 %** (1)	**3 %** (2)	**0 %** (0)

The column sums may add up to a number above 100% because in some cases two errors occurred at the same time (see Sect. 4.4). However, these combinations occurred in an unsystematic way and are therefore not reported here but can be seen in Online Resource 4. Errors marked in blue cannot occur when reading from visualizations. [a]Not applicable solution committed the conditional error, which means that participants entered the joint information at a position in the visualization that is reserved for conditional information (confirming H2a). This is particularly evident with the probability double tree, where in half of the responses, the joint probability (which cannot be entered) was confused with a conditional probability. Therefore, the reason for the large performance difference between probability double trees and natural frequency double trees could be that many participants entered the given joint probability in the wrong place instead of checking whether entering a joint probability in a double tree was possible. For the frequency 2×2 table, the most common errors were to state erroneously that joint frequencies could not be entered and/or to enter only an incomplete set of frequencies. For the frequency double tree, although there was no confusion of joint with conditional information, the joint frequencies were entered instead as marginal frequencies in 18% of the responses, which is quite remarkable in that this *marginal error* has never been reported for interpreting visualizations.

6 Discussion

Previous studies on the effect of visualizations in probabilistic situations with two binary events on participants' performance almost exclusively focused on *completely filled* visualizations, whereas entering statistical information is also an important step in solving probability tasks in school. Therefore, in this study, we theoretically embedded current research on conditional probabilities in the modeling cycle and empirically examined one step within this modeling cycle: In order to find out how well students use visualizations as a heuristic tool in probabilistic tasks with two binary events, we pre-relieved the creation of the empty structures and focused instead on students' ability and typical problems in filling in statistical information in different visualizations. In doing so, we also moved away from typical Bayesian tasks. Instead, we considered situations with two binary events in general and analyzed students' performance and difficulties with respect to entering conditional *and* joint information.

The following important results emerged: Participants' performance, as well as typical error patterns, in entering conditional or joint information depend on both the information format and the given visualization. The famous frequency effect (Gigerenzer & Hoffrage, 1995; McDowell & Jacobs, 2017), which also holds for *determining* conditional information by using completely filled visualizations (Binder et al., 2015), can furthermore be observed by *entering* frequencies into all empty visualizations except the double tree. Using natural frequencies instead of probabilities drastically reduced the already known joint occurrence error (Gigerenzer & Hoffrage, 1995; Woike et al., 2023) for the 2×2 table and the net diagram. The joint occurrence error that often appears when interpreting probability 2×2 tables and searching for conditional information almost completely disappears with frequency 2×2 tables. This error also appears when participants enter conditional information.

However, the well-known frequency effect cannot be generalized to include entering joint information. This is in line with the results of Binder et al. (2020) and Stegmüller et al. (2024) for completely filled visualizations. Thus, similar effects can be observed for *entering* joint information as for *interpreting* visualizations. Again, the most common error was to confuse joint with conditional probabilities, occurring often when only conditional probabilities could be entered (double tree) and less often when both probabilities could be entered (net diagram). Using natural frequencies made the conditional error disappear for the net diagram and the double tree.

Furthermore, in line with Cox (1997), we were also able to discover error patterns that had not yet been observed from interpreting visualizations: When students are asked to enter natural frequencies, sometimes only *one* absolute frequency is entered. Conversely, sometimes information is entered into visualizations multiple times, which also reflects a certain degree of uncertainty on the part of the students in filling in the visualizations. This degree of uncertainty is also evident when one considers how many students incorrectly checked that it was not possible to enter a particular piece of information even though it was entirely possible to do so. A previously unknown error in the frequency double tree was the confusion of joint information with marginal frequencies.

Overall, the results suggest that students lack conceptual knowledge on the different types of probabilities and/or procedural knowledge (type 1) on visualizations. This is especially apparent in the high proportion of confusions between joint probabilities and conditional probabilities in the net diagram that displays both types of probabilities simultaneously. In this context, it is also noticeable: While 93% of the students were able to correctly enter the joint probability in 2×2 tables, 84% of the students also entered the conditional probability in exactly the same place. However, the success of entering the information into the respective visualizations also depends on the given heuristic tool (i.e., the empty visualization) and the given information format. Depending on the heuristic tool selected, students with the same level of prior knowledge are likely to have varying difficulties, which must be taken into account in lessons. So, if students correctly entered a joint probability in a 2×2 table, this does not necessarily mean that students have sufficient conceptual knowledge to distinguish between joint and conditional information (Dröse et al., 2022; Shaughnessy, 1992).

6.1 Limitations

The task the participants had to solve typically does not appear in the isolated way it was examined in this study but rather is a single step in the solving process of a probabilistic task. In order to ensure that the participants used the intended visualizations and to rule out any potential difficulties in drawing the visualizations, the visualizations were already provided in the tasks. Therefore, the results relate more to the participants' conceptual knowledge on probabilities and procedural knowledge type 1 on visualizations and can only be interpreted indirectly from the perspective of visualizations as heuristic tools for modeling and problem-solving. However, it remains unknown if the errors emerged predominantly due to a lack of conceptual knowledge of the different types of probabilities, of the procedural knowledge (type 1) of visualizations (e.g., inserting a correctly decoded probability into the wrong place), or on the translation between the verbal and the mathematical representation (e.g., correctly entering an incorrectly decoded probability). Additionally, we do not know how confident participants were in their answers and if or how often they hesitated to state that information cannot be entered and applied some sort of availability heuristic instead ("just write the information in the place that looks most like what is being searched for").

The study was conducted during the pandemic of COVID-19. For this reason, participants were recruited from a university as well as from schools. Schools, in particular, were heavily affected by lockdowns and school closures, so the knowledge and performance in mathematics of the participating school students may differ from the regular level. However, previous studies (and also our data) show that effects of information format and visualization hardly differ for university and school students' Bayesian reasoning (see McDowell & Jacobs, 2017). Due to the sample size of 172 students, the results are only meaningful to a limited extent. However, even with this sample size, significant effects are already evident. In addition, the sample is a convenience sample and therefore a sample bias may have occurred (e.g., students with a particularly high or low level of conceptual knowledge in probabilities). Nevertheless, there is no reason to assume that the effects of the information format or the visualizations would substantially change with larger sample sizes or other samples—but rather that the entries into the visualizations would be better or worse in general. Furthermore, participants were only recruited in Bavaria, Germany, so a critical examination should be carried out if the results also apply to other countries or even other states in Germany since the states have different curricula.

Finally, the comparison of the different visualizations is impaired by the fact that all participants were familiar with 2×2 tables and tree-diagrams but not with double trees and net diagrams. However, double trees and net diagrams are branching structures and therefore follow the same rules for reading and entering values as well-known (single) tree diagrams. In addition, performance with 2×2 tables was not systematically better than with the unknown visualizations.

6.2 Conclusion and further research

Our results show that performance and typical errors when entering conditional and joint information depend crucially on the information format and the respective visualization, in a similar way to when reading conditional or joint information from visualizations (e.g., Binder et al., 2020). These findings can serve as important background knowledge for teachers when they work with those visualizations in class. For example, mathematics teachers can thus—depending on the visualization used—deliberately provoke typical mistakes in order to address them constructively in the classroom. Moreover, it became once more apparent that students lack conceptual knowledge of probabilities and/or procedural knowledge on visualizations. This suggests that the different functions visualizations serve cannot be treated separately from each other. Using visualizations as heuristic tools requires procedural knowledge of visualizations themselves but also conceptual knowledge on probabilities in general, that, in turn, can be fostered by using visualizations as teaching material.

In this paper, we studied students' ability to enter given values into empty visualizations primarily from the perspective of using visualizations as heuristic tools to solve probabilistic tasks. However, the ability to correctly enter numerical values into a visualization can be useful in almost every step of the modeling cycle (see Fig. 2). For example, one could use a visualization to structure the real problem situation (steps 1 and 2) or to interpret or validate the results (steps 5 and 6) that were obtained without the help of visualizations (e.g., by only using Bayes' formula). Moreover, a visualization can be drawn to inform a patient comprehensively about the result of a medical test (step 7), which also refers to the communicative function of visualizations.

Considering the overarching modeling task, future research could focus on the remaining steps of the solving process in order to discover further sources of error and better understand participants' difficulties in solving probabilistic tasks with two binary events. This also includes analyzing whether students are able to select suitable visualizations depending on the task (and the given parameters in the task) in terms of flexible use of the visualization or making connections between different representations (Acevedo Nistal et al., 2009; Novick & Hmelo, 1994; Zahner & Corter, 2010).

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10649-024-10372-y.

Acknowledgements The authors would like to thank the editor and the reviewers for their valuable comments on earlier versions of the article, Frances Lorie for editing the manuscript, Lou Sommer for the help in error analysis, and all participating students for their contribution.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

ACARA. (2022). *F-10 curriculum: Numeracy, statistics and probability*. Retrieved April 5, 2023 from https://v9.australiancurriculum.edu.au/f-10-curriculum/general-capabilities/numeracy?element=2&sub-element=0

Acevedo Nistal, A., Dooren, W., Clarebout, G., Elen, J., & Verschaffel, L. (2009). Conceptualising, investigating and stimulating representational flexibility in mathematical problem solving and learning: A critical review. *ZDM-Mathematics Education, 41*(5), 627–636. https://doi.org/10.1007/s11858-009-0189-1

Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology, 35*(5), 303–314. https://doi.org/10.1037/0022-3514.35.5.303

Awuah, F. K., & Ogbonnaya, U. I. (2020). Grade 12 students' proficiency in solving probability problems involving contingency tables and tree diagrams. *International Journal of Instruction, 13*(2), 819–834. https://doi.org/10.29333/iji.2020.13255a

Barbey, A. K., & Sloman, S. A. (2007). Base-rate respect: From ecological rationality to dual processes. *Behavioral and Brain Sciences, 30*(3), 241–54. https://doi.org/10.1017/s0140525x07001653. discussion 255-97.

Batanero, C., Estepa, A., Godino, J. D., & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education, 27*(2), 151–169. https://doi.org/10.5951/jresematheduc.27.2.0151

Biehler, R., Frischemeier, D., & Podworny, S. (2017). Reasoning about models and modelling in the context of informal statistical inference. *Statistics Education Research Journal*, *16*(2), 8–12. Retreived July 17, 2023 from https://iase-web.org/ojs/serj/article/download/593/441

Binder, K., Krauss, S., & Bruckmaier, G. (2015). Effects of visualizing statistical information - An empirical study on tree diagrams and 2 × 2 tables. *Frontiers in Psychology, 6*, 1186. https://doi.org/10.3389/fpsyg.2015.01186

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

Blum, W., & Leiss, D. (2007). *Investigating quality mathematics teaching: The DISUM project*. Retrieved September 19, 2023 http://matematikdidaktik.org/wp-content/uploads/2021/07/05_madif5.pdf#page=8

Bobek, E., & Corter, J. E. (2010). Effects of problem difficulty and student expertise on the utility of provided diagrams in probability problem solving. *Proceedings of the Annual Meeting of the Cognitve Science Society*(32). Retrieved July 3, 2023 from https://escholarship.org/content/qt4zt846s2/qt4zt846s2.pdf

Böcherer-Linder, K., & Eichler, A. (2019). How to improve performance in Bayesian inference tasks: A comparison of five visualizations. *Frontiers in Psychology, 10*, 267. https://doi.org/10.3389/fpsyg.2019.00267

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 in Psychology, 10*, 632. https://doi.org/10.3389/fpsyg.2019.00632

Büchter, T., Eichler, A., Steib, N., Binder, K., Böcherer-Linder, K., Krauss, S., & Vogel, M. (2022). How to train novices in Bayesian reasoning. *Mathematics, 10*(9), 1558. https://doi.org/10.3390/math10091558

Casscells, W., Schoenberger, A., & Graboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *The New England Journal of Medicine, 299*(18), 999–1001. https://doi.org/10.1056/NEJM197811022991808

Cox, R. (1997). Representation interpretation versus representation construction: A controlled study using switchERII. In B. Du Boulay & R. Mizoguchi (Eds.), *Frontiers in artificial intelligence and applications: vol. 39. Artificial intelligence in education: Knowledge and media in learning systems : Proceedings of AI-ED 97 world conference on artificial intelligence in education* (pp. 434–444). IOS.

Dale, D. (2020). *Fact check: Digging into Trump's false claim that CDC found that 85% of people who wear masks get the coronavirus.* CNN. Retrieved March 28, 2023 from https://edition.cnn.com/2020/10/16/politics/fact-check-trump-cdc-masks-85-percent/index.html

Díaz, C., & Batanero, C. (2009). University students' knowledge and biases in conditional probability reasoning. *International Electronic Journal of Mathematics Education, 4*(3), 131–162. https://doi.org/10.29333/iejme/234

Dröse, J., Griese, B., & Wessel, L. (2022). Prospective teachers' diagnostic judgments on students' understanding of conditional probabilities. In *Twelfth Congress of the European Society for Research in Mathematics Education (CERME12)*. Retrieved May 7, 2023 from https://hal.science/hal-03744585/

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

Eichler, A., & Böcherer-Linder, K. (2018). Categorizing errors in Bayesian situations. *Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10) Looking Back, Looking Forward*. Retrieved July, 2023 from https://iase-web.org/icots/10/proceedings/pdfs/icots10_6b1.pdf

Eichler, A., & Vogel, M. (2015). Teaching risk in school. *The Mathematics Enthusiast, 12*(1–3), 168–183. https://doi.org/10.54870/1551-3440.1342

Enaganti, I., Ganesh, N., & Mishra, B. (2022). *Inventions of interventions: Data driven strategies in pandemic research and control* (Vol. 22). National Academy of Inventors. https://doi.org/10.21300/22.2.2021.12

Falk, R. (1986). Conditional probabilities: Insights and difficulties. *Proceedings of the Second International Conference on Teaching Statistics*, 292–297. Retrieved July 10, 2023 from https://www.stat.auckland.ac.nz/~iase/publications/icots2/falk.pdf

Feufel, M. A., Keller, N., Kendel, F., & Spies, C. D. (2023). Boosting for insight and/or boosting for agency? How to maximize accurate test interpretation with natural frequencies. *BMC Medical Education, 23*(1), 75. https://doi.org/10.1186/s12909-023-04025-6

Friederichs, H., Ligges, S., & Weissenstein, A. (2014). Using tree diagrams without numerical values in addition to relative numbers improves students' numeracy skills: A randomized study in medical education. *Medical Decision Making : An International Journal of the Society for Medical Decision Making, 34*(2), 253–257. https://doi.org/10.1177/0272989X13504499

Gage, J. (2012). Towards a new probability curriculum for secondary schools. *12th International Congress on Mathematical Education*(Vol. 8).

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

Hembree, R. (1992). Experiments and relational studies in problem solving: A meta-analysis. *Journal for Research in Mathematics Education, 23*(3), 242. https://doi.org/10.2307/749120

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. 1–28). Erlbaum.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2000). Medicine. Communicating statistical information. *Science, 290*(5500), 2261–2262. https://doi.org/10.1126/science.290.5500.2261

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

Kaiser, G., & Sriraman, B. (2006). A global survey of international perspectives on modelling in mathematics education. *ZDM-Mathematics Education, 38*(3), 302–310. https://doi.org/10.1007/BF02652813

Khan, A., Breslav, S., Glueck, M., & Hornbæk, K. (2015). Benefits of visualization in the mammography problem. *International Journal of Human-Computer Studies, 83*, 94–113. https://doi.org/10.1016/j.ijhcs.2015.07.001

Kultusministerkonferenz (2022). Bildungsstandards für das Fach Mathematik: Erster Schulabschluss (ESA) und mittlerer Schulabschluss (MSA). *KMK*.

Lamanna, L., Gea, M. M., & Batanero, C. (2022). Do secondary school students' strategies in solving permutation and combination problems change with instruction? *Canadian Journal of Science, Mathematics and Technology Education, 22*(3), 602–616. https://doi.org/10.1007/s42330-022-00228-z

Maier, H., & Schweiger, F. (1999). *Mathematik und Sprache. Zum Verstehen und Verwenden von Fachsprache im Mathematikunterricht*. oebv & hpt.

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. Retrieved May 10, 2024 from https://www.ssoar.info/ssoar/bitstream/handle/document/39517/ssoar-2014-mayring-qualitative_content_analysis_theoretical_foundation.pdf

McDowell, M., & Jacobs, P. (2017). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin, 143*(12), 1273–1312. https://doi.org/10.1037/bul0000126

McDowell, M., Galesic, M., & Gigerenzer, G. (2018). Natural frequencies do foster public understanding of medical tests: Comment on Pighin, Gonzalez, Savadori, and Girotto (2016). *Medical Decision Making : An International Journal of the Society for Medical Decision Making, 38*(3), 390–399. https://doi.org/10.1177/0272989X18754508

NCTM. (2023). *Standards & positions: Data analysis and probability*. Retrieved April 5, 2023 from https://www.nctm.org/Standards-and-Positions/Principles-and-Standards/Data-Analysis-and-Probability/

Novick, Hmelo, C. E. (1994). Transferring symbolic representations across nonisomorphic problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1296–1321. https://doi.org/10.1037/0278-7393.20.6.1296

Nunes, T., Bryant, P., Evans, D., Gottardis, L., & Terlektsi, M.-E. (2014). The cognitive demands of understanding the sample space. *ZDM-Mathematics Education, 46*(3), 437–448. https://doi.org/10.1007/s11858-014-0581-3

Operskalski, J. T., & Barbey, A. K. (2016). Medicine. *Risk Literacy in Medical Decision-Making. Science, 352*(6284), 413–414. https://doi.org/10.1126/science.aaf7966

Pfannkuch, M., Ben-Zvi, D., & Budgett, S. (2018). Innovations in statistical modeling to connect data, chance and context. *ZDM-Mathematics Education, 50*(7), 1113–1123. https://doi.org/10.1007/s11858-018-0989-2

Post, M., & Prediger, S. (2022). Teaching practices for unfolding information and connecting multiple representations: The case of conditional probability information. *Mathematics Education Research Journal*, pp. 1–33. https://doi.org/10.1007/s13394-022-00431-z

Prediger, S., & Schink, A. (2009). "Three eights of which whole?" - Dealing with changing referent wholes as a key to the part-of-part-model for the multiplication of fractions. In M. Tzekaki, M. Kaldrimidou, & C. Sakonidis (Chairs), 3*3rd Conference of the International Group for the Psychology of Mathematics Education*, pp. 49–416. Thessaloniki, Greece.

Rellensmann, J., Schukajlow, S., & Leopold, C. (2017). Make a drawing. Effects of strategic knowledge, drawing accuracy, and type of drawing on students' mathematical modelling performance. *Educational Studies in Mathematics, 95*(1), 53–78. https://doi.org/10.1007/s10649-016-9736-1

Rellensmann, J. (2019). *Selbst erstellte Skizzen beim mathematischen Modellieren: Ergebnisse einer empirischen Untersuchung*. Springer.

Riccomini, P. J. (2005). Identification and remediation of systematic error patterns in subtraction. *Learning Disability Quarterly, 28*(3), 233–242. https://doi.org/10.2307/1593661

Roca, A. E., & Batanero, C. (2006). *Computing probabilities from two way tables:* An exploratory study with future teachers. *Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS7).* Retrieved July 11, 2023 from http://www.stat.auckland.ac.nz/~iase/publications/17/c413.pdf

Sáenz, C. (2009). The role of contextual, conceptual and procedural knowledge in activating mathematical competencies (PISA). *Educational Studies in Mathematics, 71*(2), 123–143. https://doi.org/10.1007/s10649-008-9167-8

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

Shaughnessy, J. M. (1992). Research on probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning (*pp. 465–494). Macmillan.

Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future. *Science, 333*(6048), 1393–1400. https://doi.org/10.1126/science.1191181
Steckelberg, A., Balgenorth, A., Berger, J., & Mühlhauser, I. (2004). Explaining computation of predictive values: 2 × 2 table versus frequency tree. A randomized controlled trial [ISRCTN74278823]. *BMC Medical Education, 4*(1), 13. https://doi.org/10.1186/1472-6920-4-13
Stegmüller, N., Binder, K., & Krauss, S. (2024). How general is the natural frequency effect? *Frontiers in Psychology.* Advance online publication. https://doi.org/10.3389/fpsyg.2024.1296359
Steib, N., Büchter, T., Binder, K., Eichler, A., Böcherer-Linder, K., Krauss, S., Vogel, M., & Sven Hilbert (2024). How to teach Bayesian reasoning. *Learning and Instruction*. Advance Online Publication.https://doi.org/10.1016/j.learninstruc.2024.102032
Uesaka, Y., Manalo, E., & Ichikawa, S. (2007). What kinds of perceptions and daily learning behaviors promote students' use of diagrams in mathematics problem solving? *Learning and Instruction, 17*(3), 322–335. https://doi.org/10.1016/j.learninstruc.2007.02.006
Wiesner, P., Binder, K., Eberl, A., & Krauss, S. (in press). Warum "20% von 50%" 10 Prozent, aber auch 40 Prozent sein können. *Stochastik in der Schule.*
Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review / Revue Internationale De Statistique, 67*(3), 223. https://doi.org/10.2307/1403699
Woike, J. K., Hertwig, R., & Gigerenzer, G. (2023). Heterogeneity of rules in Bayesian reasoning: A toolbox analysis. *Cognitive Psychology, 143*, 101564. https://doi.org/10.1016/j.cogpsych.2023.101564
Zahner, D., & Corter, J. E. (2010). The process of probability problem solving: Use of external visual representations. *Mathematical Thinking and Learning, 12*(2), 177–204. https://doi.org/10.1080/10986061003654240
Zapata-Cardona, L. (2018). Students' construction and use of statistical models: A socio-critical perspective. *ZDM-Mathematics Education, 50*(7), 1213–1222. https://doi.org/10.1007/s11858-018-0967-8
Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: The role of representation in mental computation. *Cognition, 98*(3), 287–308. https://doi.org/10.1016/j.cognition.2004.12.003

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.