



Covariational reasoning in Bayesian situations

Theresa Büchter¹ · Andreas Eichler¹ · Katharina Böcherer-Linder² ·
Markus Vogel³ · Karin Binder⁴ · Stefan Krauss⁵ · Nicole Steib⁵

Accepted: 21 October 2023 / Published online: 27 January 2024
© The Author(s) 2024

Abstract

Previous studies on Bayesian situations, in which probabilistic information is used to update the probability of a hypothesis, have often focused on the calculation of a posterior probability. We argue that for an in-depth understanding of Bayesian situations, it is (apart from mere calculation) also necessary to be able to evaluate the effect of *changes of parameters* in the Bayesian situation and the consequences, e.g., for the posterior probability. Thus, by understanding Bayes' formula as a function, the concept of covariation is introduced as an extension of conventional Bayesian reasoning, and *covariational reasoning* in Bayesian situations is studied. Prospective teachers ($N=173$) for primary ($N=112$) and secondary ($N=61$) school from two German universities participated in the study and reasoned about covariation in Bayesian situations. In a mixed-methods approach, firstly, the elaborateness of prospective teachers' covariational reasoning is assessed by analysing the arguments qualitatively, using an adaption of the Structure of Observed Learning Outcome (SOLO) taxonomy. Secondly, the influence of possibly supportive variables on covariational reasoning is analysed quantitatively by checking whether (i) the changed parameter in the Bayesian situation (false-positive rate, true-positive rate or base rate), (ii) the visualisation depicting the Bayesian situation (double-tree vs. unit square) or (iii) the calculation (correct or incorrect) influences the SOLO level. The results show that among these three variables, only the changed parameter seems to influence the covariational reasoning. Implications are discussed.

Keywords Covariational reasoning · Bayesian reasoning · Double-tree · Unit square · Natural frequencies · SOLO taxonomy

✉ Theresa Büchter
tbuechter@mathematik.uni-kassel.de

¹ Institute of Mathematics, University Kassel, Heinrich-Plett-Str. 40, 34132 Kassel, Germany

² Department of Mathematics Education, University of Freiburg, Freiburg, Germany

³ Institute of Mathematics, University of Education, Heidelberg, Heidelberg, Germany

⁴ Mathematical Institute, Ludwig Maximilian University, Munich, Germany

⁵ Faculty of Mathematics, University of Regensburg, Regensburg, Germany

1 Introduction

Breathalyser tests are used in police controls to identify people under the influence of alcohol. The proportion of drivers under the influence of alcohol varies according to different factors, such as the time of the day. So, the question arises: Does a variation of this proportion affect the validity of a (positive) test result and if so how?

In answering this question, it is key to study the conditional probability of being under the influence of alcohol if tested positive. This conditional probability is called *positive predictive value* (PPV) and can be calculated by *Bayes' formula*. Following Zhu & Gigerenzer (2006), we call such situations *Bayesian situations*, in which a binary information (positive vs. negative test result) is used to update the probability of a binary hypothesis (in this case, whether a person is under the influence of alcohol).

“The process of combining conditional probability information and base rate information to update a posterior probability” (Reani et al., 2018, p. 63) is called *Bayesian reasoning*. It is—sometimes with reference to conditional probabilities or Bayes' theorem—considered a central concept of probability literacy (Biehler & Burrill, 2011; Díaz & Batanero, 2009; Gal, 2005). Proven as important among disciplines such as medicine (Ashby, 2006) or law (Lindsey et al., 2003), but also for lay-people (e.g., when using medical diagnostic self-tests), Bayesian reasoning is necessary within general society (Spiegelhalter et al., 2011). Therefore, an in-depth understanding of probability and Bayesian situations is demanded in school mathematics curricula (Borovcnik, 2016; Kazak & Pratt, 2021).

Studies on Bayesian reasoning have mainly focused on the ability to *calculate* a conditional probability which we consider part of the aspect of *calculation*. However, in-depth understanding of Bayesian situations entails more than calculation alone (Borovcnik, 2012). Specifically, evaluating the “influence of variation of input parameters on the result” (Borovcnik, 2012, p. 21) is an important aspect. Thus, we propose an extension of conventional Bayesian reasoning (often measured with calculation) by the aspect which we call *covariation* (Büchter et al., 2022; Steib et al., 2023). Concerning the example above, covariation could mean that it is important to recognise that as the proportion of drivers under the influence of alcohol increases (e.g., Friday night compared to Monday morning), so too does the PPV.

Covariation as part of an extended Bayesian reasoning has hardly ever been investigated before but is well-studied in other areas of mathematics educational research, particularly in the understanding of functions. With regard to understanding functions, a person reasons “covariationally when she envisions two quantities' values varying and envisions them varying simultaneously” (Thompson & Carlson, 2017, p. 425). The current article links these two fields—Bayesian reasoning and covariational reasoning. Therefore, we understand Bayes' formula as a function with three independent variables and the formula's result as the dependent variable (see Section 2.1). Additionally, people's covariational reasoning in Bayesian situations is studied empirically by focusing on people's argumentation, about how changed parameters in a Bayesian situation influence the PPV. The results primarily inform the research on Bayesian reasoning, but may also contribute to research on (functional) covariational reasoning.

2 Theoretical background

2.1 Bayesian reasoning

In Bayesian situations, three probabilities are usually provided (Johnson & Tubau, 2015; Navarrete et al., 2015): the so-called base rate, true-positive rate and false-positive rate. In Table 1, we define these probabilities with references to the introductory context, in which the hypothesis H of being under the influence of alcohol is evaluated based on a positive test result (I), and provide authentic probabilities (see Ashdown et al., 2014; Lipari et al., 2017).

The PPV, $P(H|I)$, i.e., the probability that a positively tested person is *actually* under the influence of alcohol, can be calculated with Bayes' formula, $P(H|I) = \frac{P(I|H) \cdot P(H)}{P(I|H) \cdot P(H) + P(I|\bar{H}) \cdot P(\bar{H})}$ resulting in a 17% probability ($\frac{0.1 \cdot 0.9}{0.1 \cdot 0.9 + 0.9 \cdot 0.5}$). For calculation, a meta-analysis on 35 experimental studies showed a low performance of 4%, if the given information is provided in probabilities (McDowell & Jacobs, 2017). This is concerning, as performance is unsatisfyingly poor even in groups of experts who require Bayesian reasoning, such as medical practitioners (Hoffrage & Gigerenzer, 1998) and legal experts (Lindsey et al., 2003). One reason for the weak performance of calculation is base rate neglect (Kahneman & Tversky, 1982), by which people tend to overlook the influence of the base rate. Another reason is revealed by research into misleading strategies for calculation (e.g., Eichler et al., 2020; Zhu & Gigerenzer, 2006): people struggle to identify the correct sets and subsets in the complex nested-sets structure (Sloman et al., 2003) of a Bayesian situation for calculating a PPV. Rushdi & Serag (2020) pointed out that a Bayesian situation consists of 16 probabilities represented by sets and subsets, namely four single-event probabilities such as $P(H)$, eight conditional probabilities such as $P(I|H)$, and four conjunctive probabilities such as $P(H \cap I)$. For clarifying this nested-sets structure, research has identified two helpful strategies for calculation:

1. *Natural frequencies* as the format of the given statistical information.
2. *Visualisations* as a tool to structure the Bayesian situation.

Using natural frequencies was introduced by Gigerenzer & Hoffrage (1995) as the presentation of the given statistical information in pairs of natural numbers, which may represent an expected frequency (Krauss et al., 2020). In Table 1, the statistical information in the breathalyser context is presented in probabilities and natural frequencies. The meta-analysis of McDowell & Jacobs (2017) yielded that the supportive effect of natural frequencies increases performance from 4% with probabilities to 24% with statistical information in natural frequencies. Natural frequencies seemingly facilitate conventional Bayesian reasoning due to the reduced complexity of Bayes' formula for calculating the PPV (Johnson & Tubau, 2015) and due to their resemblance of a palpable situation including meaningful natural numbers instead of single-event probabilities (Todd & Gigerenzer, 2012). Moreover, the so-called nested-sets account proposes that natural frequencies more transparently represent the structure of sets and subsets in a Bayesian situation (Sloman et al., 2003).

The second helpful strategy concerns the visualisation of Bayesian situations. However, the supportive effect of this strategy is less consistently reported than the use of natural frequencies. For instance, Euler diagrams are not very supportive (Brase, 2008; Micallef et al., 2012), and a well-known tree diagram is more supportive with natural frequencies

Table 1 Bayesian situation about breath analyser tests with probabilities and natural frequencies

	Definition	Probabilities	Natural frequencies
Base rate	The a priori probability for a hypothesis, $P(H)$.	There is a 10% probability that a person who participates in a breathalyser test is under the influence of alcohol.	100 out of 1000 people who participate in a breathalyser test are under the influence of alcohol.
True-positive rate	The probability for an indicator, given a hypothesis, $P(I H)$.	If a person is under the influence of alcohol, then the probability is 90% that this person tests positive.	90 out of 100 people who are under the influence of alcohol test positive.
False-positive rate	Probability for an indicator even though a hypothesis is not actually given, $P(I \bar{H})$.	If a person is not under the influence of alcohol, then the probability is 50% that this person tests positive nevertheless.	450 out of 900 people who are not under the influence of alcohol test positive nevertheless.

than with probabilities (Binder et al., 2015). Nevertheless, regular tree diagrams are still less likely to increase the performance of calculation than other visual forms such as a *double-tree* or a *unit square*. With these latter two, the performance of calculation increases to 50–60% (Binder et al., 2020; Böcherer-Linder & Eichler, 2019). Other visualisations, such as icon arrays or 2 × 2 tables, may be even more supportive for the calculation (Böcherer-Linder & Eichler, 2019). However, neither of these visualisations can (easily) represent changes, as the given parameters are not visualised or hundreds of icons would change (Büchter et al., 2022). Yet, in double-trees and unit squares, changes can be easily visualised (see Figs. 1 and 2). Both types of visualisation have previously proven equally helpful for identifying the appropriate sets and subsets in a Bayesian situation (Böcherer-Linder & Eichler, 2019) and, thus, for calculation. Therefore, they are both used here to illustrate *situational characteristics* of covariation in Bayesian situations (understood as changes in one of the given parameters and its covarying probabilities). These can be illustrated by arrows in the double-tree or adjusted proportions in the unit square. Covariational reasoning refers to the person-specific cognitive processes involved in assessing these changes.

Covariational reasoning requires imagining how changes in the probabilities $P(H)$, $P(I|H)$ and $P(I|\bar{H})$ affect the PPV as $\frac{P(I|H)}{P(I)} = \frac{P(I|H)}{P(I|H)+P(I|\bar{H})} = \frac{P(I|H) \cdot P(H)}{P(I|H) \cdot P(H) + P(I|\bar{H}) \cdot P(\bar{H})}$ which is visualised in Figs. 1 and 2 with the relevant nodes and areas of a double-tree and a unit square, respectively, in a so-called “visual fraction” (Eichler & Vogel, 2010). Evaluating, how $P(H \cap I)$ and $P(H \cap \bar{I})$ are affected by changes of the given parameters is sufficient. Henceforth, these probabilities are referred to as the *relevant* probabilities.

The following differences between both visualisations are noticeable: The area proportionality in the unit square allows investigation into changes of relevant probabilities by (mentally) moving one of the dividing lines (which each corresponds to one given

Original Bayesian situation	False-positive rate decreases	True-positive rate decreases	Base rate increases	
				Double tree
$\frac{90}{540} = \frac{90}{90+450}$	$\frac{90}{540} = \frac{90}{90+450}$	$\frac{90}{540} = \frac{90}{90+450}$	$\frac{90}{540} = \frac{90}{90+450}$	Visual fraction (PPV)
	The frequency in the numerator, stays the same, the frequency in the denominator decreases, therefore, the fraction (PPV) increases.	The frequency in the numerator decreases and in the denominator, the frequency in the first summand decreases while the frequency in the second summand stays the same. Therefore, the relative decrease is stronger in the numerator than in the denominator, thus, the fraction (PPV) decreases.	The frequency in the numerator increases, and in the denominator, the frequency in the first summand increases while the frequency in the second summand decreases. Taken as a whole, it is unclear whether the denominator in- or decreases. However, in any case it cannot increase as much (relatively and absolutely) as the numerator, thus, the fraction (PPV) increases.	Change of fraction (PPV)

Fig. 1 Visualisation of covariation of the relevant quantities in the double-tree. Note: The false-positive rate, $P(I|\bar{H}) = 50\%$; true-positive rate, $P(I|H) = 90\%$ and base rate, $P(H) = 10\%$ are given in the double-tree. The derived natural frequencies for calculating the PPV are given in a visual fraction. The given parameters are coloured, and the influences of their change on the PPV are visually outlined by arrows

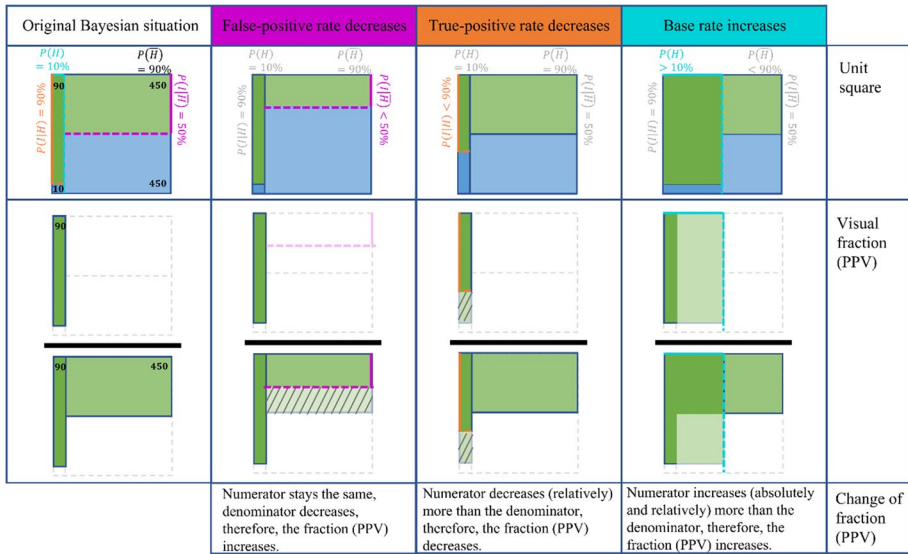


Fig. 2 Visualisation of covariation of the relevant quantities in the unit square. Note: The false-positive rate, $P(I|\bar{H}) = 50\%$; true-positive rate, $P(I|H) = 90\%$ and base rate, $P(H) = 10\%$ are given in a unit square. The derived natural frequencies and corresponding areas for calculating the PPV are given in a visual fraction. The given parameters are coloured, and the influences of their change on the PPV are visually outlined by changes in the areas

parameter). Conversely, double-trees represent the events as natural frequencies in nodes connected by branches. Therefore, with double-trees, a less visual-based but more schematic approach seems necessary for linking changes of the percentages on the branches with changes of the relevant probabilities and the PPV. Area proportionality may be particularly important for analysing variations, as changes are not only represented descriptively (i.e., numerically) but also depictively (Schnotz, 2014), i.e., through the changing areas. Additionally, for Bayesian reasoning, area proportionality has been interpreted as supportive (e.g., Micallef et al., 2012; Talbot & Schneider, 2017; Tsai et al., 2011). This may be most evident in evaluating the effect of changes of the base rate on $P(I)$, which is possible with the unit square but not with the double-tree (column 4, Figs. 1 and 2). Consequently, we expect more elaborative covariational reasoning with the unit square than with the double-tree.

Furthermore, differences in the reasoning about the different parameters (false-positive, true-positive and base rate) can be identified: When the false-positive rate changes, only one of the two relevant probabilities changes, i.e., $P(\bar{H} \cap I)$, and only the denominator of the fraction of the PPV is affected (column 2, Figs. 1 and 2). With changes in the true-positive rate, only one parameter changes, but the numerator and the denominator of the fraction of the PPV are affected (column 3, Figs. 1 and 2). Most changes have to be considered for variations of the base rate (column 4, Figs. 1 and 2): both relevant probabilities change, and both numerator and denominator are affected. Consequently, the relevant changes increase from variations of the false-positive to the true-positive to the base rate, and, therefore, cognitive load (Sweller, 2011) should be highest for changes in the base rate and lowest for changes in the false-positive rate. Consequently, we expect covariational reasoning to be least elaborate

for changes in the base rate and most elaborate when reasoning about changes in the false-positive rate. To our knowledge, only Böcherer-Linder et al. (2017) have tested covariational reasoning of Bayesian situations referring to a covariation of the base rate and the PPV with four single-choice questions. The results showed that covariational reasoning was better with unit squares than with regular tree diagrams.

2.2 Covariational reasoning outside of Bayesian reasoning

As pointed out by Zieffler & Garfield (2009), covariational reasoning has been studied in psychology, statistics education and mathematics education. Psychological research often studies prior beliefs and covariational reasoning, and research in statistics education focuses on the association between two variables based on data represented, e.g., in contingency tables or scatterplots (e.g., Batanero et al., 1997; Konold, 2002; Miguel et al., 2019). In Bayesian situations, the association of the hypothesis and information is a precondition for meaningful inferences. Thus, we are particularly interested in a functional covariation of different parameters in the Bayesian situation. Consequently, we focus on previous research in which covariational reasoning forms a crucial aspect for understanding functions.

Oerthmann et al. (2008) suggested that developing a robust understanding of functions entails “a conception that begins with a view of function as an entity that accepts input and produces output, and progresses to a conception that enables reasoning about dynamic mathematical content and scientific contexts” (p. 28). The view of a function as an “input-output-generator” is also referred to as the “mapping” aspect (Lichti & Roth, 2019). The understanding of covariation is considered more elaborate than the mapping aspect and is even viewed as the most important meaning of functions (Thompson & Harel, 2021).

We rely on Thompson & Carlson (2017) to understand a function covariationally when variations of two quantities are conceived simultaneously. Covariational reasoning refers to the cognitive processes of envisioning simultaneous variations of two quantities. Like Castillo-Garsow et al. (2013) and de Beer et al. (2015), Thompson and Carlson additionally differentiate between covariational reasoning based on thinking either in discrete points (chunky) or as a continuous process (smooth); they regard smooth covariational reasoning as more elaborate than the chunky approach.

A theoretical framework which is often used to classify students’ covariational reasoning includes the five mental actions (MA) by Carlson et al. (2002) (compare Table 2).

Previous studies highlighted that students’ activities often stagnate on the second or third mental action (e.g., Carlson et al., 2002; Fuad et al., 2019; Johnson, 2012). However, Moore & Carlson (2012) showed that students who were able to (mentally) structure the changing quantities (e.g., with an adequate sketch of the situation) were more successful to reason covariationally about the volume and height of a box. Moreover, Moss & London McNab (2011) have argued that, even for young learners, the visual and geometric representation was effective in developing a covariational understanding of linear functions (cf. also Vogel et al., 2007). Although there are also inconsistent results on the use of representations, as some imply that the graphical construction does not improve covariational reasoning (Wilkie, 2020), we consider visual representations to be supportive for covariational reasoning in Bayesian situations.

Table 2 Mental actions of covariational reasoning by Carlson et al. (2002)

Mental action	Description
1	Coordinating the value of one variable with changes in the other
2	Coordinating the direction of change of one variable with changes in the other variable
3	Coordinating the amount of change of one variable with changes in the other variable
4	Coordinating the average rate of change of function with uniform increments of change in the input variable
5	Coordinating the instantaneous rate of change of the function with continuous changes in independent variable for the entire domain of the function

2.3 Covariational reasoning and Bayesian reasoning

Studying covariational reasoning in Bayesian situations entails a few peculiarities. Firstly, we understand covariational reasoning as a higher-order ability of an extended Bayesian reasoning (compared to calculation). In a Bayesian situation, the quantity of a given parameter (e.g., base rate) and of an output (e.g., PPV) covary only indirectly through other relevant quantities, e.g., $P(H \cap I)$. For this reason, it is necessary to be aware which parameters are necessary for calculating the PPV before making judgements about its variation. As pointed out earlier, research yielded that it is an obstacle to identify the correct probabilities or sets and subsets as an input of a function and process them correctly (in Bayes' formula) for generating an output (e.g., PPV). This is different from situations such as the "bottle-problem" (Carlson et al., 2002), where "the height as a function of water that's in the bottle" (p. 360) make both the input and output explicit. For this reason, we expect more elaborate covariational reasoning when calculation is correct (as, in that case, the correct sets and subsets have been identified). Additionally, it therefore makes sense to rely on double-trees and unit squares for studying covariation in Bayesian situations because they are supportive for identifying the correct nested-sets structure of a Bayesian situation (Böcherer-Linder & Eichler, 2019; Sloman et al., 2003). Moreover, the presented research about covariational reasoning implies that a unit square with its area proportionality potentially engenders smooth covariational reasoning (if people use the structure to imagine smooth changes in the areas and different sets and subsets), whereas working with double-trees is more likely to engender a chunky covariational reasoning (if people imagine changes with this structure as discrete changes in the natural frequencies).

Secondly, quantities are a key element in the theory of covariational reasoning of functions (Thompson & Carlson, 2017). By "quantity", Thompson & Carlson (2017) refer to a conceptualisation of an object's attribute which can be measured. Quantities in Bayesian situations may be probabilities that can be measured in a frequentist interpretation as a relative frequency in a long run of repetitions. Probabilities thus represent an estimation of a future relative frequency. Therefore, conceiving a probability as a quantity may entail imagining that an event, such as testing positive, can (in a long run of repetitions of testing people) be measured by the relative frequency of positively tested people among the entire test sample. Consequently, conceptualising quantities in a Bayesian situation which is characterised by probabilities, may be more challenging than in a situation which is characterised by natural frequencies, since probabilities require a frequentist interpretation for conceiving quantities.

Thirdly, often functions refer to physical quantities where the direction of change is easily understandable, such as the "bottle-problem" (e.g., Carlson et al., 2002; Johnson, 2015), in which the volume and height of a water bottle are studied simultaneously. Of course, the volume increases with an increase of the height. Thus, a mental action on the third level by

Carlson et al. (2002; cf. Table 2) seems possible. In Bayesian situations, probabilities as quantities are not observable unlike the volume or height of a bottle. Accordingly, the effect of one quantity on a second is not as easily understandable as in such physical problems, and therefore reasonings on level 3 seem beyond an expectable covariational reasoning in Bayesian situations. For instance, the influence of changes of the base rate on the PPV is not at all obvious, as previous research has repeatedly pointed out for the calculation but not yet for covariation (Kahneman & Tversky, 1982; Stengård et al., 2022).

3 Research questions and hypotheses

The current study aims to investigate how people reason about covariation in Bayesian situations, hence to measure covariational reasoning in Bayesian situations. The research questions (RQ), with hypotheses (H), are summarised in Table 3.

4 Methods

4.1 Participants

In our study, 173 prospective teachers (136 female, 35 male, 2 unknown) for primary school (112 participants) or secondary school (61 participants) from two German universities (82 at university 1, 91 at university 2) participated voluntarily, and written

Table 3 Research questions with hypotheses

Research question	Hypothesis
RQ1: Does the specific parameter (false-positive rate, true-positive rate, or base rate), which is varied, affect the elaborateness of covariational reasoning?	H1a: We expect the covariational reasoning about changes to the false-positive rate to be more elaborate than that about changes to the true-positive rate. H1b: We expect the covariational reasoning about changes to the true-positive rate to be more elaborate than about changes to the base rate (see Section 2.1).
RQ2: Does correct calculation affect the elaborateness of the covariational reasoning in the Bayesian situation?	H2: Correct calculation is based on correctly identifying the relevant probabilities or, sets and subsets in the nested-sets structure of a Bayesian situation. For this reason, we expect more elaborate covariational reasoning for subjects who calculate the PPV correctly, because then all parameters relevant for the covariation of the given and requested quantity have been identified (Section 2.1).
RQ3: Does the visualisation (double-tree or unit square), which is used to represent the Bayesian situation, affect the elaborateness of the covariational reasoning?	H3: Area proportionality of the unit square potentially engenders smooth covariational reasoning while working with a double-tree more likely engenders chunky covariational reasoning (Section 2.1). Smooth covariational reasoning is regarded as more elaborate (Section 2.2). Moreover, the area proportionality of the unit square could be of particular support for covariational reasoning as it helps to observe how the relevant subsets change, especially for base rate changes (Section 2.1). Thus, we expect more elaborate covariational reasoning with unit squares compared to double-trees, particularly for base rate changes.

informed consent was obtained. The 173 prospective teachers are part of a larger sample ($N=229$) that was tested beyond covariational reasoning as we operationalise it here. In this paper, we refer to the sub-sample of 173 prospective teachers, because the remaining 56 participants received test items whose responses are not appropriate for a systematisation of covariational reasoning as developed for this paper. From all participants, 143 were in the first to third semesters, and 30 participants were in the fourth or higher semesters. The study programmes for all students include lectures and seminars in mathematics and mathematics education. Only the study programmes for future *secondary* school teachers include compulsory courses on probability, generally in the fourth semester. Thus, 26 out of the 173 students should have been enrolled on a probability course (without a specific focus on Bayesian reasoning) during or prior to participation. These 26 students showed no differences from the others regarding their (i) knowledge of the visualisation, (ii) calculation, (iii) selected single-choice answer about direction of change, and (iv) elaborateness of the given reasonings (chi-square tests, Bonferroni adjusted). Conditional probabilities are part of the German school curriculum, but in school, regular tree diagrams with probabilities or 2×2 tables are generally used as visualisations. Thus, we did not expect familiarity with the visualisations used in this study (double-tree and unit square). Participants could win one of three 75€ Amazon vouchers (university 1) or were allowed to skip a mandatory coursework for participation (university 2).

4.2 Study design

The participants were randomly assigned to experimental groups. The Bayesian situation, the visualisation and the changed parameter in the situations varied between the experimental groups. About half of the participants were assigned to breathalyser tests ($N=84$) and the other half to mammography screenings ($N=89$). Because we found no differences between the elaborateness of the reasonings (for measurement of the elaborateness, see Section 4.3), we refer to the responses of both Bayesian situations without further differentiation. Based on this, of the 175 participants, $N=87$ had a double-tree and $N=86$ had a unit square to represent the Bayesian situation. In both visualisation groups, parts of the groups worked with one changed parameter (false-positive rate, true-positive rate *or* base rate). The sizes of the related sub-samples are given in Fig. 3.

The study was carried out as an online survey. The participants received a short introduction to the visualisation used later, as no familiarity was expected and to ensure comparable familiarity with both visualisations (see supplementary material, <https://osf.io/fbdn2/>). Then, they were asked to calculate the PPV in a Bayesian situation to which statistical information was provided in the visualisation of their introductory example (Fig. 3, Table 4). Next, a variation of the Bayesian situation was described, where the false-positive rate, the true-positive rate *or* the base rate was changed (base rate decreased, others increased). Next, the students were asked, in a single-choice question, how this change affected the PPV (decreases, stays the same, increases) and then to provide reasoning for their selected answer. This design is displayed in Fig. 3 and the wording of the tasks in Table 4. According to Table 4, the calculation task required the numerator and denominator as absolute frequencies (in a whole number). As all participants entered absolute frequencies in these two fields, it is likely that they conceived the sets which they entered as possible to measure, which is essential for covariational reasoning (see above).

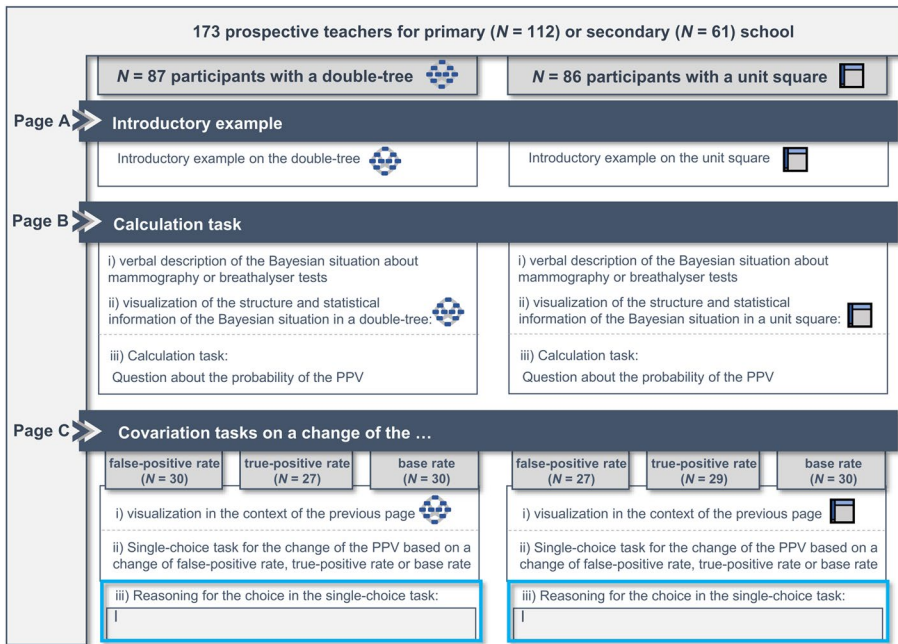


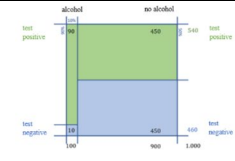
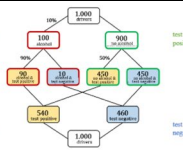
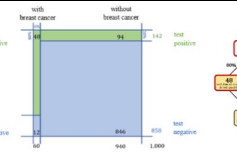
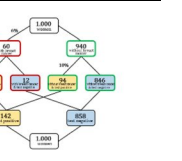
Fig. 3 Design of the study. Note: This figure primarily specifies the sample structure as well as the digital layout and order of the tasks (reasoning tasks highlighted, as they are the primary data which are analysed in this contribution). For the wording of the tasks, see Table 4

4.3 Analysis

We first analyse our data qualitatively to structure the reasonings into less and more elaborate categories, and then quantitatively by comparing the frequencies in the derived categories between experimental groups. Therefore, according to Kansteiner & König (2020), our study can be understood as mixed-methods research.

To qualitatively analyse our data, we used the Structure of Observed Learning Outcome (SOLO) taxonomy, which provides a classification “for discriminating between responses of different qualities” (Biggs & Collis, 1982, p. 17). The SOLO taxonomy is frequently used in statistics education, when deductive categories of observable elaborateness of statistical reasoning are investigated (e.g., Eichler & Vogel, 2012; Watson et al., 2003). Five levels are differentiated in the SOLO taxonomy, which differ in terms of the amount of relevant information and their interrelation. Covariational reasoning in a Bayesian situation entails identifying the relevant sets and subsets ($H \cap I$, $\bar{H} \cap I$ and I), identifying changes in the relevant quantities, $P(H \cap I)$ and $P(\bar{H} \cap I)$ or $P(I)$, and analysing how these changes affect the PPV (compare Section 2.1). We further argued above that this identification of the relevant sets and subsets is peculiar to covariational reasoning in Bayesian situations. Therefore, in comparison to other existing frameworks (e.g., the mental actions according to Carlson et al., 2002), the SOLO taxonomy in this case seems particularly fitting, as it can be judged by how far students are able to identify and covary the different *relevant* (sub)-sets. Thus, arguing with irrelevant quantities is treated as prestructural reasoning. Furthermore, since we included the variation of one quantity (false-positive rate,

Table 4 Wording of the different tasks and contexts

		Page B (Calculation task)			
		Breathalyser test (BT)	Mammography screenings (MS)		
Context with visualization		In Regensburg, 1000 drivers were given a breathalyser test to check their intoxication in traffic controls last August. In Regensburg, only a small proportion of the drivers was under the influence of alcohol. The Dräger-6510 test has the following characteristics: the majority of people who are under the influence of alcohol is detected by the breathalyser test, and therefore tests positive. A large proportion of the people who are not under the influence of alcohol still test positive. <ul style="list-style-type: none"> • Unit square (with statistical information) • Double-tree (with statistical information) 	In Hessen every year about 1000 women participate in a mammography-screening; these present no symptoms for breast cancer and their near relatives are not known to have had breast cancer. Of these women, only a small proportion actually has breast cancer. In the mammography screenings, a large proportion of the women with breast cancer is detected and therefore tests positive. A small proportion of the women without breast cancer is falsely tested positive.		
	Calculation task	How likely is a person to actually be under the influence of alcohol if she tests positive in the breathalyser test? In order to calculate this probability, you need to form a fraction (numerator/denominator). Please determine:	How likely is a woman to actually have breast cancer if she tests positive in the mammography screening?		
		Numerator (as a whole number): _____ Denominator (as a whole number): _____ Probability (in percent with two decimals): _____ %			
		Page C (Covariation tasks)			
Visualization of the previous page					
	Single-choice task	<p>False-positive rate</p> <p>Consider the probability that a person, who is not under the influence of alcohol, falsely tests positive is larger than 50%. The other values are the same as in the visualisation. (BT)</p>	<p>True-positive rate</p> <p>Consider the probability that a person, who is under the influence of alcohol, actually tests positive is larger than 90%. The other values are the same as in the visualisation. (BT)</p>	<p>Base rate</p> <p>Consider the probability that a woman with breast cancer actually tests positive is larger than 80%. The other values are the same as in the visualisation. (MS)</p>	<p>Base rate</p> <p>Consider the probability that a person is under influence of alcohol is 6%. The other values are the same as in the visualisation. (BT)</p>
Reasoning		How does that change the probability that a person is actually under the influence of alcohol, if she tests positive (compared to the original situation in the visualisation)? The probability ...			
		<input type="radio"/> decreases <input type="radio"/> stays the same <input type="radio"/> increases Please provide reasoning for your selected answer:			

true-positive rate, base rate) in the task and asked for the simultaneous variation of the PPV, every answer referring to the variation of any other relevant quantity is treated as *covariational reasoning*. Based on the SOLO taxonomy, we have developed a schema for the elaborateness of covariational reasoning in Bayesian situations (Table 5), which are assigned according to a coding system supplied in the supplementary material (<https://osf>.

Table 5 Descriptions of the SOLO levels

SOLO level	Description of the level according to Biggs & Collis (1982)	Description of the level for covariational reasoning in a Bayesian situation
<i>Prestructural</i>	Irrelevant data is used in the response; the response closes without even seeing the problem; cue and response are confused.	As a result of changing one quantity (false-positive rate, true-positive rate, base rate), only effects on irrelevant quantities are described; incorrect changes of relevant quantities are described; direct link from the changed parameter to the change of the PPV is described.
<i>Unistructural</i>	Only one relevant data is linked with the cue; response closes too quickly.	As a result of changing one quantity (false-positive rate, true-positive rate, base rate), the direction of change of one relevant quantity ($P(H \cap I)$, $P(\bar{H} \cap I)$ or $P(I)$) is correctly described (+incorrect changes of further relevant quantities are possibly given).
<i>Multistructural</i>	Isolated relevant cues are linked to the cue; response does not relate the relevant cues to one another.	As a result of changing one quantity (false-positive rate, true-positive rate, base rate), directions of changes of at least two relevant quantities ($P(H \cap I)$, $P(\bar{H} \cap I)$ or $P(I)$) are correctly described, without relating these changes correctly to the direction of change of the PPV.
<i>Relational</i>	The relevant cues are related to another and linked to the cue.	As a result of changing one quantity (false-positive rate, true-positive rate, base rate), directions of changes of both relevant quantities ($P(H \cap I)$ and $P(\bar{H} \cap I)$), or $P(H \cap I)$ and $P(I)$) are correctly described, and these changes are correctly related to the direction of change of the PPV.
<i>Extended abstract</i>	In addition to relational level, generalisations are provided to the given specific situation.	As a result of changing one quantity (false-positive rate, true-positive rate, base rate) and additionally to the observations necessary for the relational level, the strictly monotonous relation between the changed quantity and the PPV is identified and therefore the described changes are generalised to any given Bayesian situation.

In this table, both the original taxonomy (middle column) and their application to covariational reasoning in Bayesian situations (right column) are listed

[io/fbdn2/](#)). The inter-rater reliabilities of two coders for all categories are reported as Cohen's kappas in the coding system (all above 0.73 after a training phase).

To identify statistically significant predictors for the elaborateness of covariational reasoning, we used a multinomial logistic regression for predicting the SOLO level (Field et al., 2012). As predictors, we rely on the changed parameter (false-positive, true-positive or base rate) in a first regression model. We also rely on the calculation (people's ability to compute the PPV correctly) and the visualisation (double-tree or unit square) in two further regression models. Our data satisfies the assumptions for multinomial logistic regressions (Field et al., 2012): the independence of errors is based on our between-subject design. As the predictors in the regression models cannot correlate due to the dummy coding of each predictor that signifies a different experimental group, multicollinearity is not given. Finally, a linear relationship between any continuous predictor and the logit of the outcome variable is not applicable, since we only use dummy-coded predictors. Data analyses were carried out with R in version 4.3.0 and the package

“mlogit.” The data and the script of the analyses can be found in the supplementary material (<https://osf.io/fbdn2/>).

5 Results

5.1 Qualitative analysis of covariational reasoning in Bayesian situations

The reasonings were assigned to four of the five SOLO levels. No reasonings were observed in the extended abstract level. For all other levels, we provide one example for each of the changed parameters in Table 6. As we focus on a probability which the participant addressed, we do not further differentiate between the different ways to represent a probability, for example as a percentage, a fraction or as natural frequencies (cf. Section 2.1).

Reasonings without mentioning any quantities (examples 1, 6 and 11) were assigned to “non-codable.” In examples 2, 7 and 12, no change is described on either of the relevant quantities: example 2 only states that sensitivity and base rate (both irrelevant) remain constant; example 7 falsely argues that the sensitivity equals the PPV (incorrect); example 12 only repeats that the base rate decreases, without reaching any conclusions about the relevant quantities. Thus, examples 2, 7 and 12 belong to the prestructural level. However, examples 3, 8 and 13 covary the changed parameter (provided in the instruction) with one other relevant quantity: example 3 correctly describes the change of number for all positively tested; example 8 correctly observes that $H \cap I$ increases; example 13 correctly recognises that with a decrease of the base rate (value of 0.1), the probability $P(H \cap I)$ with the value of 9% decreases as well. Yet, these responses fail to address the change of the other relevant quantities. Hence, examples 3, 8 and 13 link only one relevant data with the cue and belong to the unistructural level. In examples 4, 9 and 14, the consequences on both relevant quantities (the true *and* false-positives) are correctly identified. Still, these observations do not correctly describe how this affects the PPV: example 4 incorrectly assumes that the PPV remains unchanged; example 9 does not make any explicit inferences on the PPV; example 14 incorrectly describes the effect on the PPV, as the denominator actually decreases. Therefore, these reasonings belong to the multistructural level. Finally, examples 5, 10 and 15 correctly identify the relevant changes and also correctly interrelate them verbally (example 15) or by means of a calculation (examples 5 and 10).

In total, 173 participants provided reasonings for changes in the false-positive (57), true-positive (56) or base rate (60) which were coded into the different SOLO levels (Table 7).

The proportion of reasonings with a specific SOLO level among each parameter is represented in Fig. 4.

5.2 Quantitative analysis of RQ1: elaborateness of covariational reasoning in Bayesian situations with different parameter changes

Descriptively, the proportion of reasonings, which is non-codable, does not differ substantially between the different parameter changes (compare Fig. 4). However, the results suggest that from false-positive to true-positive to base rate, the proportion of prestructural level increases, whereas the proportion of the multistructural and relational level decreases. This may indicate a confirmation of our hypotheses that reasoning about the false-positive rate should be easier than reasoning about the true-positive rate (H1a) and that reasoning

Table 6 Examples of reasonings

Changed parameter	Example (SOLO level)	Reasoning (translated into English from German)
Increase of the false-positive rate	Example 1 (non-codable)	Denominator increases, result of the fraction decreases.
	Example 2 (prestructural)	Because the probability that a person under influence of alcohol is tested positive has nothing to do with the probability of the people who are not under the influence of alcohol. The probability for people under the influence of alcohol remains the same (90%, 10%).
	Example 3 (unistructural)	Number of all positively tested increases.
	Example 4 (multistructural)	One probability does not influence the other. Even if more people are falsely tested positive, the number of people who are under the influence of alcohol and test positive remains the same.
	Example 5 (relational)	The number of positively tested increases, the 90 stays the same. Let's assume that not 450 are falsely tested positive but 700, then this means 90/790 are actually positive and that is smaller than 90/540.
Increase of the true-positive rate	Example 6 (non-codable)	Then the tests are more precise.
	Example 7 (prestructural)	A person under the influence of alcohol who actually tests positive = a person who is actually under the influence of alcohol if she tests positive.
	Example 8 (unistructural)	It increases, as there are more drivers who are under the influence of alcohol and also test positive.
	Example 9 (multistructural)	The number of true-positive tests increases, while the number of false-positive tests remains the same.
	Example 10 (relational)	Original situation is $90/540=16.7\%$. Now, the same number is added to numerator and denominator: $95/545=17.34\%$. The number increases, as the ratio changes
Decrease of the base rate	Example 11 (non-codable)	Because the probability is also smaller.
	Example 12 (prestructural)	Previously, the probability that a person is under the influence of alcohol was exactly 10%. If this proportion decreases, the probability decreases as well.
	Example 13 (unistructural)	$0.1 \times 0.9 = 9\%$; smaller than $0.1 \times 0.9 =$ smaller than 9%
	Example 14 (multistructural)	Frequency of people under the influence of alcohol smaller, 90% of them positive \rightarrow fewer under the influence of alcohol and tested positive and number of false positively tested larger, denominator larger and numerator smaller, thus result is smaller.
	Example 15 (relational)	If the probability decreases that a person is under the influence of alcohol, then there is also a decrease in the number of people who are under the influence of alcohol, and consequently also in the number of people who are under the influence of alcohol and test positive. Simultaneously the probability increases that a person is not under the influence of alcohol, and the number of people who are not under the influence of alcohol and falsely test positive increases likewise. Therefore, the proportion of people who are actually under the influence of alcohol, among all people who test positive, decreases and therefore also the probability that a person is actually under the influence of alcohol, who tests positive.

Table 7 Absolute frequencies of SOLO levels for different parameter changes

SOLO level	Change in:			Sum
	False-positive rate	True-positive rate	Base rate	
Absolute frequency of SOLO level (proportion among the changed parameter)				
Non-codable	20 (0.35)	21 (0.38)	20 (0.33)	61
Prestructural	7 (0.12)	13 (0.23)	26 (0.43)	46
Unistructural	20 (0.35)	20 (0.36)	11 (0.18)	51
Multistructural	5 (0.09)	1 (0.02)	2 (0.03)	8
Relational	5 (0.09)	1 (0.02)	1 (0.02)	7
Extended abstract	0 (0.0)	0 (0.0)	0 (0.0)	0
Sum	57	56	60	173

Proportion of SOLO level among each changed parameter (false-positive, true-positive, or base rate) is given in brackets

about the base rate should be even more difficult than reasoning about the true-positive rate (H1b). To test the statistical significance, we carried out a multinomial logistic regression with reasonings about the true-positive rate as the reference group. The base line of the SOLO level is set to the prestructural level. In all following analyses, those reasonings which are “non-codable” are excluded. Also, reasonings in the multistructural and relational levels are very rare (compare Table 7) and structurally very similar (e.g., both correctly describe changes in both relevant quantities and only differ with regard to the consequence on the PPV). Thus, for the statistical analyses, they are combined in one level.

Table 8 shows the results of the multinomial logistic regression for predicting the different SOLO levels. False-positive rate and base rate are predictors, which are compared to the changes of the true-positive rate in the reference group. The intercepts show how much more likely a specific SOLO level is in the reference group. For example, the unistructural level is 1.54 times as likely as the prestructural level in reasonings about the true-positive rate (calculation of odds: $\frac{\text{observations unistructural level}}{\text{observations prestructural level}} = \frac{20}{13} \approx 1.54$, compare Tables 7 and 8). Yet, the regression coefficient ($b = 0.43, p = 0.23$), and thus the difference between pre- and

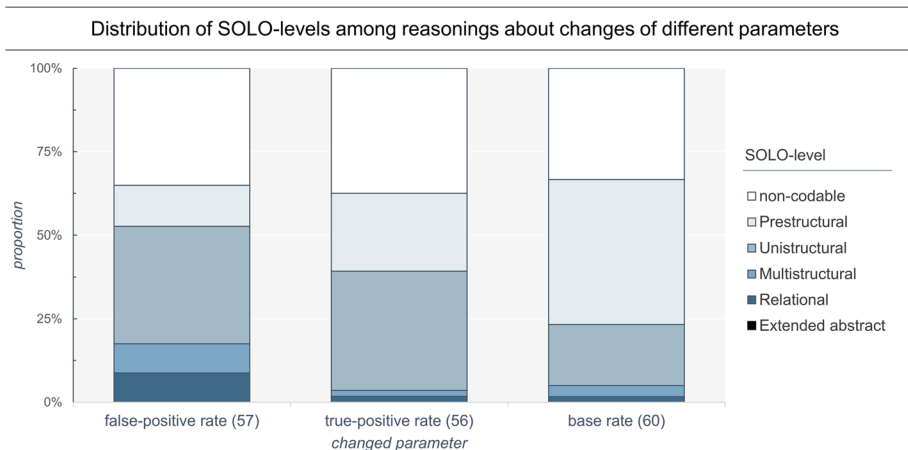


Fig. 4 Proportions of the different SOLO levels among changes of the different parameters

Table 8 Results of the multinomial logistic regression for predicting the elaborateness of the reasoning based on the changed parameter

	<i>b</i> (<i>SE</i>)	<i>p</i>	95% <i>CI</i> for odds ratio		
			<i>Lower</i>	<i>Odds ratio</i>	<i>Upper</i>
Prestructural vs. unistructural					
Intercept	0.43 (0.36)	0.23	0.77	1.54	3.09
False-positive rate	0.62 (0.57)	0.27	0.61	1.86	5.63
Base rate	-1.29 (0.51)	0.01*	0.10	0.28	0.74
Prestructural vs. (multistructural or relational)					
Intercept	-1.87 (0.76)	0.01*	0.04	0.15	3.09
False-positive rate	2.23 (0.91)	0.01*	1.57	9.29	54.77
Base rate	-0.29 (0.97)	0.77	0.11	0.75	5.06

Significance codes: * < 0.05

The reference group of the regression are reasonings about the true-positive rate and the baseline of the outcome variable is the prestructural level. The table contains the regression coefficients (*b*) and their standard errors (*SE*), the *p*-value of the regression coefficients (*p*), as well as 95% confidence interval (*CI*) for the odds ratios

unistructural level in the reference group is not significant. However, the regression coefficient for the intercept for multistructural or relational level ($b = -1.87, p = 0.01$) is significant; thus, it is less likely for a reasoning about the true-positive rate to belong to the multistructural or relational level than the prestructural level. Knowing that a reasoning is given about the base rate (instead of the true-positive rate) decreases the chances (odds ratio: $\frac{\text{observations unistructural level base rate}}{\text{observations prestructural level base rate}} = \left(\frac{11}{26}\right) \approx 0.28$, compare Tables 7 and 8) for a reasoning to belong to the unistructural (compared to the prestructural) level ($b = -1.29, p = 0.01$). Furthermore, a reasoning about the false-positive rate is more likely than a reasoning in the reference group (about the true-positive rate) to belong to the multistructural or relational level compared to the prestructural level ($b = 2.23, p = 0.01$). Taken together, reasoning about changes in the base rate increases the chances of remaining on the lowest level (prestructural) whereas reasoning about changes in the false-positive rate increases the chances of belonging to the higher levels (multistructural or relational levels). Therefore, H1a and H1b are confirmed.

5.3 Quantitative analysis of RQ2 and RQ3: calculation and visualisation as predictors for the elaborateness of covariational reasoning

We hypothesised that the calculation (RQ2) and the visualisation (RQ3) affect the elaborateness of covariational reasoning. For testing these hypotheses, we ran two further multinomial logistic regressions. The proportions of the different SOLO levels among correct and incorrect calculation as well as in both visualizations is displayed in Fig. 5.

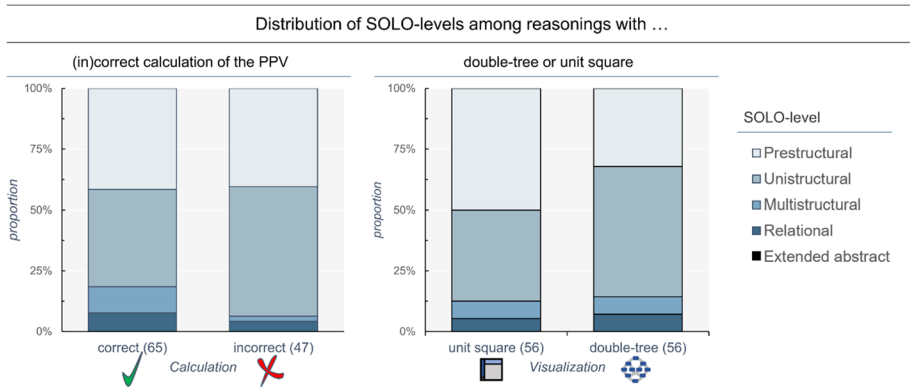


Fig. 5 Proportions of the different SOLO levels among codable reasoning with (i) an (in)correct calculation of the PPV and (ii) a unit square or a double-tree

Firstly, we tested whether calculation can predict the SOLO level of the reasoning. We set the reference group to reasonings with an incorrect calculation of the PPV and used the prestructural level as the baseline of the outcome variable. The results show that correct calculation has no significant influence on the SOLO level. However, correctly calculating the PPV descriptively increases the chances that a reasoning belongs to the multistructural or relational level ($b = 1.04$, $p = 0.15$, $OR = 2.82$) while decreasing the chances of belonging to the unistructural level as compared to the prestructural level ($b = -0.31$, $p = 0.45$, $OR = 0.73$). Consequently, H2 cannot be confirmed despite a descriptive tendency for the hypothesis.

Secondly, we tested whether the visualisation (double-tree, unit square) can predict the SOLO level of the reasoning. According to H3, the visualisation's influence should be greater for the base rate; thus, we also included the parameter as a predictor. We set the reference group to reasonings with the double-tree about the false-positive rate *or* true-positive rate and used the prestructural level as the baseline of the outcome variable. The results show that a change from the double-tree to the unit square does not significantly alter the odds of a reasoning belonging to the unistructural level ($b = -0.71$, $p = 0.20$, $OR = 0.49$) or multistructural or relational level ($b = -0.41$, $p = 0.58$, $OR = 0.67$) compared to the prestructural level for the reasonings about the false-positive and true-positive rate as well as for reasonings about the base rate ($b = -0.32$, $p = 0.73$, $OR = 0.72$ for the unistructural level and $b = -0.76$, $p = 0.61$, $OR = 0.47$ for the multistructural or relational level). Therefore, H3 cannot be confirmed.

6 Discussion

In the context of functions, reasoning about judgements of covariation has been shown to reveal understanding of covariation (e.g., Saffran et al., 2019). We applied this idea to study covariational reasoning in Bayesian situations by analysing *how* people *reason* about covariation in such situations. With this approach, new perspectives can be gained focusing primarily on Bayesian reasoning and, secondarily, about covariational reasoning.

The results of our study revealed that elaborateness of covariational reasoning differs and can be categorised according to the SOLO taxonomy. With this categorisation, it is

possible to identify qualitative differences between the reasonings (e.g., outlining irrelevant changes vs. focusing only on isolated (relevant) changes of the Bayesian situation). Secondly, variables which affect the elaborateness of the covariational reasoning have been identified by a quantitative analysis of the distribution of the SOLO levels. Thereby, the changed parameter seems to be most influential. In accordance with our hypotheses, reasonings about the consequence of changes in the false-positive rate are more elaborate than reasonings about the consequences of changes in the true-positive rate. Furthermore, reasonings about the consequences of changes in the base rate are the least elaborate. The correct calculation in the Bayesian situation does not significantly predict the covariational reasoning even though it descriptively increases the chances of a reasoning belonging to the multistructural or relational level. The visualisation had no significant influence on the covariational reasoning.

6.1 Implications about the association between calculation and covariational reasoning

Bayesian reasoning has previously been studied with a focus on calculation (McDowell & Jacobs, 2017). Using calculation in the form of a chunky covariational reasoning (de Beer et al., 2015) would have been possible, if the participants re-calculated the PPV with self-set new values but was only observed in few reasonings (e.g., example 10 in Table 6). Additionally, our results indicate that correct calculation is not sufficient to predict covariational reasoning and, therefore, covariation as part of an extended Bayesian reasoning may be a (partially) distinct aspect from conventional Bayesian reasoning (often measured by calculation). This would be in contrast to empirical results about functional thinking, where Lichti & Roth (2019) could not observe covariation as a separate dimension. This may be particularly surprising as, in our study, design calculation was tested before covariational reasoning, differently from other studies on functional covariational reasoning (e.g., Carlson et al., 2002), and could therefore have triggered a stronger connection between both aspects. Moreover, this result could be surprising from the perspective of Bayesian situations that necessitate the identification of relevant probabilities for calculating a PPV as well as for covariational reasoning. Yet, another interpretation is possible. The results suggest that calculation does not affect the unistructural level, and this could be in line with research on typical erroneous strategies in Bayesian reasoning (Binder et al., 2020; Eichler et al., 2020): Erroneous strategies for calculation often entail indicating one of the relevant quantities as the PPV (e.g., joint-occurrence strategy, which means erroneously identifying $P(H \cap I)$ as PPV), or combining one relevant quantity with an irrelevant quantity (e.g., Pre-Bayes strategy, which means erroneously identifying $\frac{P(H)}{P(I)}$ as PPV). If these erroneous strategies are used for calculation, they may still lead to somewhat elaborate covariational reasonings, as changes in at least some of the relevant quantities are considered. This may suggest that correct calculation is primarily important for higher level reasonings, as only in the multistructural and relational level are *both* relevant quantities addressed. This is in line with the descriptively large difference in the proportion of reasonings in the multistructural and relational level between reasonings with correct and incorrect calculation. However, also a small proportion of multistructural and relational reasonings occur, even if the PPV was calculated incorrectly. These reasonings concern changes in the false-positive rate, which may imply that elaborating changes of the false-positive rate prompts a switch in the calculation strategy. In conclusion, further

studies should analyse the consistency of the calculation strategy in tasks about covariation and identify whether covariation is a separate dimension of Bayesian reasoning.

6.2 Implications about the effect of the visualisation

Visualisation did not affect covariational reasoning. This is remarkable, as the only previous study on covariation in Bayesian situations by Böcherer-Linder et al. (2017) showed an advantage of the unit square over the *regular* tree diagram for covariational reasoning. The authors interpreted this as a superiority of the area proportionality of the unit square. An alternative interpretation is possible: the regular tree diagram is already less supportive for calculation compared with the unit square (Böcherer-Linder & Eichler, 2017, 2019), thus the difference concerning covariational reasoning might have been based on the difference in the performance of calculation. Another explanation may be that the area proportionality of the unit square (possibly linked to a smooth covariational reasoning) is not actually any more supportive than the node-branch structure immanent in both regular and double-tree diagrams (possibly linked to a chunky covariational reasoning) for covariational reasoning. Finally, a difference of familiarity may have influenced the results about the visualisation. Even though both visualisations are unfamiliar, the node-branch structure of a double-tree diagram and its relation to the well-known probability tree diagrams may induce less alienation than a unit square with the unknown area proportionality. Therefore, the participants may not have been capable to use the advantages of the unit square (Section 2.1) as they were not acquainted enough with it. Consequently, in our larger research project, Train-Bayes (http://bayesianreasoning.de/en/bayes_en.html), we study covariational reasoning after participating in a training course on Bayesian reasoning with the double-tree or the unit square.

6.3 Implications about the influence of the changed parameter

For changes in the base rate, it was particularly harder than for changes of the true- (or false-) positive rate to identify consequences on even one relevant quantity. Hence, many reasonings about base rate changes stagnate on the lowest level. This confirms previous results about Bayesian reasoning, as it has often been assumed that struggles with calculation are based on the base rate neglect (Kahneman & Tversky, 1982; Stengård et al., 2022). Furthermore, an analysis of reasonings in the unistructural level reveals that: mentioning the consequences on the quantity of the true-positives $P(H \cap I)$ is most likely for changes in the true-positive (75%) and base rate (64%). Hence, potentially, the quantity of the false-positives is more easily overlooked than the quantity of the true-positives. This would be in line with research about erroneous strategies of Bayesian reasoning, as no erroneous strategy is known, which focuses on the quantity of the false-positives (Binder et al., 2020; Eichler et al., 2020). However, among reasonings in the unistructural level about changes in the false-positive rate, only 20% mention the true-positives, 30% mention the quantity of all positives and 50% the quantity of the false-positives. Thus, considering changes in the false-positive rate may accentuate the relevance of the quantity of false-positives, $P(\bar{H} \cap I)$. Consequently, considering changes in the false-positive rate may also increase the performance of calculation, as it may clarify the identification of the false-positives as a relevant quantity for calculating the PPV.

Finally, with the prompted questions, we have induced reasonings on mental actions 1 and 2 (Carlson et al., 2002) only. Even though previous studies on covariational reasoning have illustrated that students often stagnate on mental action 3 (Fuad et al., 2019; Johnson, 2012), we have shown that even among mental actions 1 and 2 the reasonings are far from elaborate, with only a small minority reaching levels above the unistructural level. This may be based on the peculiarity in the current study of examining covariational reasoning with Bayes' formula as the underlying function. This implies that it is necessary to identify the relevant quantities in a Bayesian situation—or, respectively, the relevant sets and subsets in the nested-sets structure of a Bayesian situation—not only to calculate a PPV, but also to reason about a covariation of a changed parameter in a Bayesian situation and a simultaneous change of the PPV.

6.4 Limitations

Our sample consisted of future mathematics teachers and for this reason, it could be analysed if the results can be transferred to another population. Furthermore, our data are reasonings for a single-choice answer, which were categorised according to the SOLO taxonomy. Our results imply that the SOLO levels correlate strongly with answers in the single-choice format, which was the assessment method used in the only known previous study about covariational reasoning in Bayesian situations by Böcherer-Linder et al. (2017). Still, we are aware that our results are limited to this assessment method (for alternative assessments see Steib et al., 2023). Moreover, the reasonings were provided by the participants without an opportunity to ask for clarification. Consequently, a considerable amount of data (about 35%) was not able to be coded with our coding scheme (category “non-codable”). Additionally, some answers were possibly coded to a lower level even though a participant actually had a more elaborate understanding. For instance, sometimes participants may not have been able to (correctly) *verbally* express their understanding of the situation, as this is challenging (Díaz & Batanero, 2009; Post & Prediger, 2020). One person wrote about an increase in the false-positive rate: “the quantity of positively tested increases. That means that, in total, the proportion of people who are under the influence of alcohol and test positive, decreases.” Possibly, the person confused the meaning of “and” with a conditional meaning (compare Hertwig et al., 2008). Moreover, in our analyses, we did not differentiate between varying representations which were used to refer to the probabilities (e.g., percentage vs. natural frequency). Natural frequencies may ease the concept of the quantity, making it possible to measure and therefore facilitate covariational reasoning. Thus, these should be studied in the future, to check whether natural frequencies as the format of the given information have an effect on the covariational reasoning. Concluding, the current results should be interpreted as a first approach to covariation in Bayesian situations. A consecutive interview study might provide an in-depth understanding. Finally, all differences (regarding the Bayesian situation, visualisation and parameter change) are observed between distinct subgroups. Even though we could not identify significant differences in the subgroups of the participants, future studies might need to validate whether the observed differences can also be replicated in a within-subject study design.

7 Conclusion

The results of our study provide a multi-faceted view on covariational reasoning in Bayesian situations. We demonstrated that covariational reasoning differs with regard to its elaborateness and that the parameters of the Bayesian situation seem to be more influential for the covariational reasoning than the specific visualisation or the ability to calculate the posterior probability (i.e., the PPV). All things considered, covariational reasoning is not as elaborate as the relevance of Bayesian reasoning for an in-depth understanding of probability in school (as well as for the general public) would require. Thus, we recommend further research on covariation as an aspect of extended Bayesian reasoning, in order to identify helpful strategies and the development of covariational reasoning, possibly with the help of training courses.

Author contributions All authors (TB, AE, KB-L, MV, KB, SK and NS) contributed to the study conception and design. Material preparation and data collection were performed by TB and NS. Data analysis was carried out by TB. The first draft of the manuscript was written by TB, AE, KB-L and MV. After the reviews, the manuscript was revised by TB and AE. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. The research leading to these results received funding from Deutsche Forschungsgemeinschaft (DFG) under the Grants EIC773/4-1 and KR 2032/6-1.

Data material and code availability All data as well as the R-script for the data analyses are available in the Open Science Framework repository <https://osf.io/fbdn2/>

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

- Ashby, D. (2006). Bayesian statistics in medicine: A 25 year review. *Statistics in Medicine*, 25(21), 3589–3631. <https://doi.org/10.1002/sim.2672>
- Ashdown, H. F., Fleming, S., Spencer, E. A., Thompson, M. J., & Stevens, R. J. (2014). Diagnostic accuracy study of three alcohol breathalysers marketed for sale to the public. *BMJ Open*, 4(12), e005811. <https://doi.org/10.1136/bmjopen-2014-005811>
- Batanero, C., Estepa, A., & Godino, J. D. (1997). Evolution of students understanding of statistical association in a computer-based teaching environment. In J. B. Garfield & G. Burrill (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics* (pp. 191–205). International Statistical Institute.
- de Beer, H., Gravemeijer, K., & van Eijck, M. (2015). Discrete and continuous reasoning about change in primary school classrooms. *ZDM – Mathematics Education*, 47(6), 981–996. <https://doi.org/10.1007/s11858-015-0684-5>

- Biehler, R., & Burrill, G. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *New ICMI Study Series: Vol. 14. Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study: The 18th ICMI Study* (pp. 57–70). Springer Science+Business Media B.V.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy; structure of the observed learning outcome*. Academic Press. <http://www.sciencedirect.com/science/book/9780120975525>
- 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>
- Böcherer-Linder, K., & Eichler, A. (2017). The impact of visualizing nested sets. An Empirical study on tree diagrams and unit squares. *Frontiers in Psychology*, 7, 2026. <https://doi.org/10.3389/fpsyg.2016.02026>
- 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>
- Böcherer-Linder, K., Eichler, A., & Vogel, M. (2017). The impact of visualization on flexible Bayesian reasoning. *Avances De Investigación En Educación Matemática*, 11, 25–46. <https://doi.org/10.35763/aiem.v11i1.169>
- Borovcnik, M. (2012). Multiple perspectives on the concept of conditional probability. *Avances De Investigación En Educación Matemática*, 2, 5–27. <https://doi.org/10.35763/aiem.v1i2.32>
- Borovcnik, M. (2016). Probabilistic thinking and probability literacy in the context of risk Pensamento probabilístico e alfabetização em probabilidade no contexto do risco. *Educação Matemática Pesquisa Revista do Programa de Estudos Pós-Graduados em Educação Matemática*, 18(3), 1491–1516. <https://revistas.pucsp.br/index.php/emp/article/view/31495>
- Brase, G. L. (2008). Pictorial representations in statistical reasoning. *Applied Cognitive Psychology*, 23(3), 369–381. <https://doi.org/10.1002/acp.1460>
- 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>
- Büchter, T., Steib, N., Böcherer-Linder, K., Eichler, A., Krauss, S., Binder, K., & Vogel, M. (2022). Designing visualizations for Bayesian problems according to multimedia principles. *Education Sciences*, 12(11), 739. <https://doi.org/10.3390/educsci12110739>
- Carlson, M., Jacobs, S., Coe, E., Larsen, S., & Hsu, E. (2002). Applying covariational reasoning while modeling dynamic events: A framework and a study. *Journal for Research in Mathematics Education*, 33(5), 352. <https://doi.org/10.2307/4149958>
- Castillo-Garsow, C., Johnson, H. L., & Moore, K. C. (2013). Chunky and smooth images of change. *For the Learning of Mathematics*, 33(3), 31–37. <https://flm-journal.org/Articles/11A85FC0E320DF94C00F28F2595EAF.pdf>
- 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>
- Eichler, A., Böcherer-Linder, K., & Vogel, M. (2020). Different visualizations cause different strategies when dealing with Bayesian situations. *Frontiers in Psychology*, 11, 1897. <https://doi.org/10.3389/fpsyg.2020.01897>
- Eichler, A., & Vogel, M. (2010). Die (Bild-) Formel von Bayes [The (picture-) formula of Bayes]. *PM-Praxis der Mathematik in der Schule*, 52(32), 25–30.
- Eichler, A., & Vogel, M. (2012). Basic modelling of uncertainty: Young students' mental models. *ZDM – Mathematics Education*, 44(7), 841–854. <https://doi.org/10.1007/s11858-012-0451-9>
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE.
- Fuad, Y., Ekawati, R., Sofro, A., & Fitriana, L. D. (2019). Investigating covariational reasoning: What do students show when solving mathematical problems? *Journal of Physics: Conference Series*, 1417(1), 12061. <https://doi.org/10.1088/1742-6596/1417/1/012061>
- Gal, I. (2005). Towards “probability literacy” for all citizens: Building blocks and instructional dilemmas. In G. A. Jones (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning* (pp. 39–63). Springer. https://doi.org/10.1007/0-387-24530-8_3
- 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>

- Hertwig, R., Benz, B., & Krauss, S. (2008). The conjunction fallacy and the many meanings of and. *Cognition*, 108(3), 740–753. <https://doi.org/10.1016/j.cognition.2008.06.008>
- Hoffrage, U., & Gigerenzer, G. (1998). Using natural frequencies to improve diagnostic inferences. *Academic Medicine*, 73(5), 538–540. <https://doi.org/10.1097/00001888-199805000-00024>
- Johnson, E. D., & Tubau, E. (2015). Comprehension and computation in Bayesian problem solving. *Frontiers in Psychology*, 6, 938. <https://doi.org/10.3389/fpsyg.2015.00938>
- Johnson, H. L. (2012). Reasoning about quantities involved in rate of change as varying simultaneously and independently. In R. L. Mayes & L. Hatfield (Eds.), *Quantitative Reasoning and Mathematical Modeling: A Driver for STEM Integrated Education and Teaching in Context* (pp. 39–53). University of Wyoming.
- Johnson, H. L. (2015). Secondary students' quantification of ratio and rate: A framework for reasoning about change in covarying quantities. *Mathematical Thinking & Learning*, 17(1), 64–90. <https://doi.org/10.1080/10986065.2015.981946>
- Kahneman, D., & Tversky, A. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (1st ed., pp. 153–163). Cambridge Univ. Press. <https://doi.org/10.21236/ada099501>
- Kansteiner, K., & König, S. (2020). The role(s) of qualitative content analysis in mixed methods research designs. *Forum Qualitative Sozialforschung Forum: Qualitative Social Research*, 21(1). <https://doi.org/10.17169/fqs-21.1.3412>
- Kazak, S., & Pratt, D. (2021). Developing the role of modelling in the teaching and learning of probability. *Research in Mathematics Education*, 23(2), 113–133. <https://doi.org/10.1080/14794802.2020.1802328>
- Konold, C. (2002). Teaching concepts rather than conventions. *New England Journal of Mathematics*, 34(2), 69–81.
- Krauss, S., Weber, P., Binder, K., & Bruckmaier, G. (2020). Natürliche Häufigkeiten als numerische Darstellungsart von Anteilen und Unsicherheit – Forschungsdesiderate und einige Antworten [Natural frequencies as a numerical representation of proportions and uncertainty – research desiderata and some answers]. *Journal für Mathematik-Didaktik*, 41(2), 485–521. <https://doi.org/10.1007/s13138-019-00156-w>
- Lichti, M., & Roth, J. (2019). Functional thinking—A three-dimensional construct? *Journal Für Mathematik-Didaktik*, 40(2), 169–195. <https://doi.org/10.1007/s13138-019-00141-3>
- Lindsey, S., Hertwig, R., & Gigerenzer, G. (2003). Communicating Statistical DNA Evidence. *Jurimetrics*, 43, 147–163.
- Lipari, R. N., Hughes, A., Bose, J. (2017). Driving under the influence of alcohol and illicit drugs. In *The CBHSQ Report*. Substance Abuse and Mental Health Services Administration (US), Rockville (MD).
- 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>
- Micallef, L., Dragicevic, P., & Fekete, J.-D. (2012). Assessing the effect of visualizations on Bayesian reasoning through crowdsourcing. *IEEE Transactions on Visualization & Computer Graphics*, 18(12), 2536–2545. <https://doi.org/10.1109/TVCG.2012.199>
- Miguel, M., Ernesto, S., & Eleazar, S. (2019). Covariational reasoning patterns exhibited by high school students during the implementation of a hypothetical learning trajectory. Eleventh Congress of the European Society for Research in Mathematics Education (CERME11). Utrecht University. <https://hal.science/hal-02412811/>
- Moore, K. C., & Carlson, M. P. (2012). Students' images of problem contexts when solving applied problems. *The Journal of Mathematical Behavior*, 31(1), 48–59. <https://doi.org/10.1016/j.jmathb.2011.09.001>
- Moss, J., & London McNab, S. (2011). An approach to geometric and numeric patterning that fosters second grade students' reasoning and generalizing about functions and co-variation. In J. Cai & E. J. Knuth (Eds.), *Advances in Mathematics Education. Early algebraization: A global dialogue from multiple perspectives* (pp. 277–301). Springer. https://doi.org/10.1007/978-3-642-17735-4_16
- Navarete, G., Correia, R., Sirota, M., Juanchich, M., & Huepe, D. (2015). Doctor, what does my positive test mean? From Bayesian textbook tasks to personalized risk communication. *Frontiers in Psychology*, 6, 1327. <https://doi.org/10.3389/fpsyg.2015.01327>
- Oerthmann, M., Carlson, M., & Thompson, P. W. (2008). Foundational reasoning abilities that promote coherence in student's function understanding. In M. P. Carlson & C. Rasmussen (Eds.), *Notes: v.73. Making the Connection: Research and Teaching in Undergraduate Mathematics Education* (pp. 27–42). Mathematical Association of America.
- Post, M., & Prediger, S. (2020). Decoding and discussing part-whole relationships in probability area models: The role of meaning-related language. In *Seventh ERME Topic Conference on Language in the Mathematics Classroom*. Montpellier, France. <https://hal.archives-ouvertes.fr/hal-02970604/>

- Reani, M., Davies, A., Peek, N., & Jay, C. (2018). How do people use information presentation to make decisions in Bayesian reasoning tasks? *International Journal of Human-Computer Studies*, *111*, 62–77. <https://doi.org/10.1016/j.ijhcs.2017.11.004>
- Rushdi, A. M. A., & Serag, H. A. M. (2020). Solutions of ternary problems of conditional probability with applications to mathematical epidemiology and the COVID-19 pandemic. *International Journal of Mathematical, Engineering & Management Sciences*, *5*(5), 787–811. <https://doi.org/10.33889/IJMEMS.2020.5.5.062>
- Saffran, A., Barchfeld, P., Alibali, M. W., Reiss, K., & Sodian, B. (2019). Children's interpretations of covariation data: Explanations reveal understanding of relevant comparisons. *Learning & Instruction*, *59*, 13–20. <https://doi.org/10.1016/j.learninstruc.2018.09.003>
- Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *Cambridge handbooks in psychology. The Cambridge handbook of multimedia learning* (pp. 72–103). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.006>
- Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior & Human Decision Processes*, *91*(2), 296–309. [https://doi.org/10.1016/S0749-5978\(03\)00021-9](https://doi.org/10.1016/S0749-5978(03)00021-9)
- Spiegelhalter, D., Pearson, M., & Short, I. (2011). Visualizing uncertainty about the future, 333(6048), 1393–1400. <https://doi.org/10.1126/science.1191181>
- Steib, N., Krauss, S., Binder, K., Büchter, T., Böcherer-Linder, K., Eichler, A., & Vogel, M. (2023). Measuring people's covariational reasoning in Bayesian situations. *Frontiers in Psychology*, *14*, 1184370. <https://doi.org/10.3389/fpsyg.2023.1184370>
- Stengård, E., Juslin, P., Hahn, U., & van den Berg, R. (2022). On the generality and cognitive basis of base-rate neglect. *Cognition*, *226*, 105160. <https://doi.org/10.1016/j.cognition.2022.105160>
- Sweller, J. (2011). Cognitive Load Theory. In J. P. Mestre & B. H. Ross (Eds.), *Psychology of learning and motivation: v. 55. The psychology of learning and motivation*, *55*, *Cognition in education* (pp. 37–76). Academic Press. <https://doi.org/10.1016/b978-0-12-387691-1.00002-8>
- Talbot, A. N., & Schneider, S. L. (2017). Improving accuracy on Bayesian inference problems using a brief tutorial. *Journal of Behavioral Decision Making*, *30*(2), 373–388. <https://doi.org/10.1002/bdm.1949>
- Thompson, P. W., & Carlson, M. P. (2017). Variation, covariation, and functions: Foundational ways of thinking mathematically. In J. Cai (Ed.), *Compendium for Research in Mathematics Education* (pp. 421–456). National Council of Teachers of Mathematics.
- Thompson, P. W., & Harel, G. (2021). Ideas foundational to calculus learning and their links to students' difficulties. *ZDM—Mathematics Education*, *53*(3), 507–519. <https://doi.org/10.1007/s11858-021-01270-1>
- Todd, P. M., & Gigerenzer, G. (2012). *Ecological Rationality: Intelligence in the World*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195315448.001.0001>
- Tsai, J., Miller, S., & Kirlik, A. (2011). Interactive visualizations to improve Bayesian reasoning. *Proceedings of the Human Factors & Ergonomics Society Annual Meeting*, *55*(1), 385–389. <https://doi.org/10.1177/1071181311551079>
- Vogel, M., Girwidz, R., & Engel, J. (2007). Supplantation of mental operations on graphs. *Computers & Education*, *49*(4), 1287–1298. <https://doi.org/10.1016/j.compedu.2006.02.009>
- Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science & Technology*, *34*(1), 1–29. <https://doi.org/10.1080/0020739021000018791>
- Wilkie, K. J. (2020). Investigating students' attention to covariation features of their constructed graphs in a figural pattern generalisation context. *International Journal of Science & Mathematics Education*, *18*(2), 315–336. <https://doi.org/10.1007/s10763-019-09955-6>
- 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>
- Zieffler, A. S., & Garfield, J. B. (2009). Modeling the growth of students' covariational reasoning during an introductory statistics course. *Statistics Education Research Journal*, *8*(1), 7–31. <https://doi.org/10.52041/serj.v8i1.455>